

Department of Laboratory Medicine
Clinical Research Center
Karolinska Institutet, Stockholm, Sweden

FUNCTIONAL ADAPTATION OF C-MYC AND ITS ROLE IN LYMPHOMA- ASSOCIATED GENE REGULATION

Amir Nematollahi Mahani



**Karolinska
Institutet**

Stockholm 2018

All previously published papers were reproduced with permission from the publisher.

Published by Karolinska Institutet.

Printed by EPRINT

© Amir Nematollahi Mahani, 2018

ISBN 978-91-7831-123-1

Functional adaptation of c-Myc and its role in lymphoma-associated gene regulation

THESIS FOR DOCTORAL DEGREE (Ph.D.)

By

Amir Nematollahi Mahani

Principal Supervisor:

Professor Anthony Wright
Karolinska Institutet
Department of Laboratory Medicine
Clinical Research Center

Co-supervisor(s):

Assistant Professor Samir El-Andaloussi
Karolinska Institutet
Department of Laboratory Medicine
Clinical Research Center

Opponent:

Associate Professor Mikael Lindström
Karolinska Institutet
Department of Medical Biochemistry and
Biophysics

Examination Board:

Professor Ann-Kristin Östlund Farrants
Stockholm University
Department of Department of Molecular
Biosciences
The Wenner-Gren Institute

Professor Peter Zaphiropoulos
Karolinska Institutet
Department of Biosciences and Nutrition

Professor Lars-Gunnar Larsson
Karolinska Institutet
Department of Microbiology, Tumor and Cell
Biology

To my beloved family

ABSTRACT

The Myc proto-oncogene, is a highly disordered (lack of structure formation) transcription factor (TF), which can bind to its partner proteins and regulate different biological functions of the cells including proliferation, cell cycle, differentiation and apoptosis. In general, Myc deregulation is a major prime force in human tumors, which contributes to their uncontrolled cell proliferation, metastasis and tumor cell immortalizations. In particular, Myc expression influences the transcriptional profile of cells by promoting RNA Polymerase II gene transcription to produce mRNA, as well as the transcription of the rRNA and tRNA genes transcribed by RNA Polymerase I and III, respectively. Thus, controlling expression of ribosome components, required for protein synthesis, appears to be an important role of Myc in normal and cancer cells.

In this thesis, I have studied phylogenetic and molecular evolution of the Myc family proteins, for the first time exploiting their protein order/disorder properties and the extent of their conservation through the Metazoan and beyond. We systematically analyzed the predicted protein disorder profile of Myc family proteins using a range of different algorithms. Therefore, we showed that all Myc proteins are structurally disordered TFs and most of the interaction domains of c-Myc are within disordered regions. Moreover, Using Intrinsically Disordered Protein (IDP) profiles we established a new way to evaluate the evolution of TFs based on their disorder profile. Use of IDR predictions instead of protein sequences produced a better-supported phylogenetic tree of Myc proteins, including large clades containing c-Myc, MycN, MycL and dMyc proteins. In addition, we analyzed the effect of Burkitt's lymphoma (BL) mutations on the disorder profile and suggested that these adaptive BL-Associated Mutations (BL-AM) could change the local conformation of c-Myc and thereby its functions.

Next, we studied Myc in the nucleolus, an adaptive context that has scarcely been studied, and its involvement in spatial chromatin domain organization of the rRNA genes. Accordingly, we found that Myc activation caused altered spatial organization of the mammalian rDNA by tethering the rDNA to the nucleolar scaffold/matrix via non-transcribed intergenic spacer sequences of the rDNA. In addition, in rat fibroblast cell lines we found that matrix associated rRNA genes are hypo-methylated on DNA sites in their upstream core promoter regions (CpG site at position -145).

Finally, I characterized lymphoma-associated gene expression induced by wild type Myc and how it is adapted in response to BL-AM (T58A and T58I). For this purpose, I established a cell system, consisting of low passage primary B-cells transduced with lentivirus expression constructs, in which wild type or BL-AM Myc could be induced to varying degrees by doxycycline in order to progressively promote a lymphoma-like phenotype. The transduced cells also constitutively over-expressed the BMI1 and BCLXL proteins to inhibit apoptosis. Progressive increase in Myc expression was associated with a progressive increase in cell proliferation, size and proportion of cells in S-phase for both wild type Myc and BL-AM,

albeit with some differences between the different Myc proteins. RNA-Seq was used to measure cellular transcripts at seven different doxycycline concentrations for cells expressing wild type and BL-AM Myc. Generalized linear models (as implemented in the MaSigPro package (v3.3)) were then used to identify differentially regulated genes (DEG) with regard to Myc level and/ or mutant status. Thus, we found 4443 DEG common to all three-cell system as well as 543 DEG deregulated only in T58A and T58I cells. On the other hand, the results show DEG common between wild type and T58A (n=553) as well as between wild type and T58I (n=1062). Further analysis, identified 15 gene clusters with different patterns of differential gene expression and these genes were enriched in generally distinct sets of gene ontology terms, indicating little functional overlap between clusters. The data identify gene sets induced by Myc as the cells convert to lymphoma-like cells as well as gene sets where one or both BL-AM augment changes induced by wild type Myc.

LIST OF SCIENTIFIC PAPERS

This thesis is based on the following papers

- I. **Amir Mahani**, Johan Henriksson, Anthony P. H. Wright. (2013) Origins of myc proteins - using intrinsic protein disorder to trace distant relatives. (PloS ONE, 8:e75057, 2013).
- II. Chiou-Nan Shiue, **Amir Nematollahi-Mahani** and Anthony P. H. Wright. Myc-induced anchorage of the rDNA IGS region to nucleolar matrix modulates growth-stimulated changes in higher-order rDNA architecture. (Nucleic Acids Res 42: 5505-5517, 2014)
- III. **Amir Mahani**, Gustav Arvidsson, Alf Grandien Anthony P. H. Wright. Global gene regulation changes associated with c-Myc activation and Burkitt's Lymphoma Myc mutations during conversion of B-cells to Lymphoma-like cells (Manuscript 2018)

CONTENTS

1	INTRODUCTION TO MYC AND CANCER.....	1
1.1	MYC	1
1.2	Myc structure and DNA binding sites.....	1
1.3	Myc activation	1
1.4	Myc protein and its cellular functions.....	2
1.4.1	Myc functional domains	2
1.4.2	Myc-Box I.....	2
1.4.3	Myc-Box II	2
1.4.4	bHLHZ and MIZ-1 mediate repression.....	3
1.5	Myc controlling cellular behavior	3
1.5.1	Cell cycle	3
1.5.2	Apoptosis	4
1.5.3	Transformation.....	4
1.5.4	Degradation.....	4
2	MYC EVOLUTION, SEQUENCE OR PROTEIN STRUCTURE?	7
2.1	Myc evolution and protein sequence.....	7
2.2	Intrinsic disorder profile of proteins, new insights on evolution.....	8
3	RIBOSOME BIOGENESIS	10
3.1	Mammalian nucleolus structure and rDNA	11
3.2	Nucleolar scaffold and DNA compartmentalization	11
3.3	Ribosomal RNA genes transcription.....	12
3.4	Role of Myc in rDNA transcription	13
4	ROLE OF MYC IN LYMPHOMA MALIGNANCIES.....	17
4.1	Lymphoma malignancies	17
4.1.1	T cell lymphocytes.....	17
4.1.2	B-cell lymphocytes	17
4.2	Introduction to Burkitt's Lymphoma	18
4.2.1	Myc regulation in Burkitt's Lymphoma.....	18
5	AIMS.....	26
6	MATERIAL AND METHODS	22
6.1	Lentivirus vector construction.....	22
6.2	Making the triple hit myc-driven B-cell lymphoma-like cell system.....	23
6.3	Design pipeline for read alignment and subtraction of human reads from mouse reads.....	24
6.4	Using PONDR®VLXT for prediction of BL-AM protein disorder profile.....	24
6.5	Analysis of the nuclear matrix attachment and DNA methylation assay.....	24
7	RESULTS	26
7.1	Paper I: Phylogenetic and molecular evolution studies of Myc family proteins.....	27
7.2	Paper II: Functional analysis of Myc family proteins in mammalian cells.....	28

7.3	Paper III: lymphoma-associated gene expression changes in an inducible model of Myc-driven B-cell lymphoma	29
8	DISCUSSION AND FUTURE PERSPECTIVES.....	33
9	ACKNOWLEDGEMENTS.....	37
10	REFERENCES	41

LIST OF ABBREVIATIONS

bHLHZ	Basic Helix Loop Helix leucine Zipper
BL	Burkitt's Lymphoma
BL-AM	Burkitt's Lymphoma-Associated Mutations
CTD	C-Terminal Domain
DEG	Differentially Expressed Genes
DFC	Dense Fibrillar Component
Fbw7	E3 ubiquitin ligase Fbw7
FCs	Fibrillar Center
GO	Gene Ontology
GC	Granular Component
H3, 4	Histone 3, 4
IDR	Intrinsic Disorder Region
IGS	InterGenic Spacer
MB I, II	Myc Box I, Myc Box II
NOR	Nucleolar Organization Region
rDNA	Ribosomal DNA
RNA Pol I, II, III	RNA Polymerase I, II, III
rRNA	Ribosomal RNA
S/MAR	Scaffold/Matrix Attachment Regions
Skp2	E3 ubiquitin ligase Skp2
SL1	Selective factor 1
T58A	Threonine 58 substitution to Alanine
T58I	Threonine 58 substitution with Isoleucine
TAD	Transcription Activation Domain
TFs	Transcription Factors
TRRAP	Transactivation/Transformation Associated Protein domain
UBF	Upstream Binding Factor
UCE	Upstream Core Element
3C	Chromosome Confirmation Capture

1 INTRODUCTION TO MYC AND CANCER

1.1 MYC

Myc (*v-myc*) was first identified in bird myelocytomatosis (a leukemic disorder) caused by integration of avian oncogenic retroviruses (MC29)[1]. Thereafter, different *v-myc* homologs were identified in human and characterized as a proto-oncogene family comprising *c-myc*, *N-myc* and *L-myc* genes[2, 3]. The Myc proto-oncogene family (c-Myc, N-Myc and L-Myc) has been studied for more than 35 years in cancer biology. Myc Transcription factors (TF) regulate about 15% of protein coding genes and have been implicated in regulation of different biological functions, such as cell proliferation, differentiation, metabolism and apoptosis[4-7]. In humans, the Myc protein family is associated with different types of cancers. While MycN deregulation is found in neuroblastoma and some other childhood cancers [8], MycL is involved in small lung cell cancers [8]. However, altered expression of c-Myc is frequently found in many different cancer types [8-10].

1.2 MYC STRUCTURE AND DNA BINDING SITES

Myc is a basic Helix-Loop-Helix Zipper (bHLHZ) protein, which heterodimerizes with the MAX protein (bHLHZ) to regulate its target genes[11, 12]. To function as a regulatory transcription factor, the monomers assemble on the DNA and Myc-MAX heterodimers bind to the canonical (5'-CACGTG-3') or non-canonical (5'-CANNTG-3') DNA sequences, known as E-boxes [7, 12]. Different Myc family proteins bind to the same site in DNA, while c-Myc has an omnipresent expression in different cell types, MycN and MycL are expressed during specific embryonic stages and later in specific tissues [13]. Levels of Myc expression are crucial for early embryonic development whereas, deletion of either N-Myc or c-Myc is lethal at day 9-11 in mouse embryos [14, 15].

1.3 MYC ACTIVATION

Myc expression is regulated by a wide range of mitogens, growth factors and cytokines that induce proliferation and modulate cell differentiation[16]. On the other hand Myc down regulation is associated with stimulation of differentiation or anti-proliferation signals[17]. As mentioned, Myc has a broad influence on the expression of protein coding genes (15%) transcribed by RNA Polymerase II, and many of these are involved in protein synthesis. However, increased protein synthesis requires production of all components of ribosomes including rRNA and tRNA. It is thus of interest that unlike most transcription factors, Myc is a direct regulator of tRNA and rRNA genes transcribed by RNA Polymerases I and III in addition to its regulation of protein coding genes [4, 5, 18-21]. In general, Myc can alter gene expression and histone marks of its target genes, by recruiting chromatin regulatory complexes via its N-terminal domain, including TRRAP, TBP, GCN5 and Tip60 [22-26], which locally modify histone H3 and H4 acetylation [27].

1.4 The Myc protein and its cellular functions

1.4.1 Myc functional domains

The Myc proteins consist of two main domains, the N-terminal transactivation domain (TAD) and the C-terminal domain (CTD). An 84 amino acid region of the CTD domain is a bHLHZ region which is essential for Myc heterodimerization with MAX protein and specific binding to DNA sequences, known as E-boxes (5'-CACGTG-3')(Figure 1). Mutations or the interaction with other proteins to the bHLHZ of Myc can lead to the loss of association with Max and blocking of Myc transcriptional activity[7].

1.4.2 Myc-Box I

Myc, like most transcription factors, has a conserved DNA-binding domain, which is in the CTD of Myc. However, Myc also has several evolutionary conserved regions in the N-terminal domain known as Myc homology boxes [13]. The TAD domain of Myc has two conserved regions known as Myc-Box (MB) I and MBII (Figure 1), which were earlier shown to be required for Myc induced transformation activity [28]. MBI (amino acid 45-63) is a binding site for several proteins that contribute to Myc activation and degradation (Figure 1)[6, 29]. Threonine 58 (T58) and Serine 62 (S62) are important Myc phosphorylation sites, where S62 phosphorylation stabilizes and activates Myc while phosphorylation of only T58 targets Myc degradation via Fbw7 (E3 ubiquitin ligase) (Figure 3) [30, 31]. The Myc half-life is about 24 minutes and mutations in the MBI region can affect its transformational activity by reducing its degradation [32]. Different MBI mutations have been reported in tumors, which can alter the turnover of Myc [33, 34]. In particular, the T58 amino acid substitution to alanine (T58A) and isoleucine (T58I) are reported in lymphoma tumors and their functional differences were characterized in different cell lines [32, 34, 35]. T58 mutations can alter Myc phosphorylation, Fbw7 dependent degradation and thus increase Myc half-life (T58A \approx 60 min and T58I \approx 30 min), contributing to Myc induced transformation [31, 32].

1.4.3 Myc-Box II

MBII plays a major role in Myc transcriptional activities by recruiting different co-factors such as TIP60, SKP2 and TRRAP (Figure 3)[6]. While MBII deletion (Myc Δ MBII) causes G2 arrest in cells and causes defective Myc transcriptional activity, cellular transformation and apoptosis, the mutant protein can partially induce proliferation [31, 36]. Thus, gene expression studies of Myc Δ MBII found that 90% of Myc target genes were affected by MBII deletion in comparison with WTMyC [36]. Deletion of MBII and MBI can inhibit TRRAP (TRansactivation/tRansformation Associated Protein) binding to Myc. TRRAP is an essential co-factor needed for Myc to induce transformation and transcriptional activity [23]. However, some Myc target genes are regulated independently of Myc-TRRAP binding, as shown in the Myc Δ MBII cell line [36].

1.4.4 bHLHZ and MIZ-1 mediate repression

Although Myc is known as a transcriptional activator, recent studies demonstrate that Myc can function as a transcriptional repressor by binding to Miz1 (Figure 3) and blocking MIZ-1 target gene transcription[6]. Myc-Max binds to MIZ-1 and inhibits P300 recruitment by MIZ-1 on its target genes by further recruitment of DNMT3a (DNA methyltransferase)[6, 37]. On the other hand, recent studies uncover MYC-MIZ-1 target genes independently of Myc and MIZ-1 binding sequence[37]. The same study found Myc repressed genes had E-boxes in promoters that lacked MIZ-1 binding sites but were enriched in SP1 binding site[37]. All together, these findings suggest that Myc modulating transcriptional repression is affected by Myc-MIZ-1 ratio and other protein interaction[37].

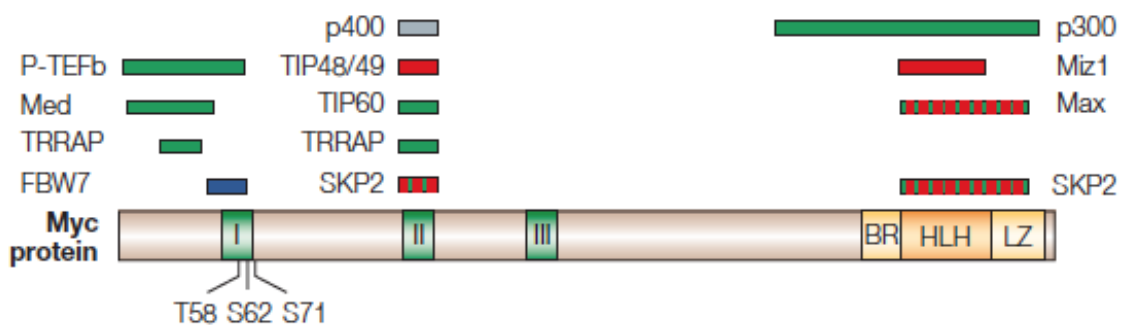


Figure 1. Domains of Myc and their binding proteins. The N terminus of Myc has three highly conserved elements, known as Mycboxes I–III. The C terminus contains the basicregion/helix–loop–helix/leucine-zipper (BR/HLH/LZ) domain. T58, S62 and T71 are known phosphorylation sites of Myc, and are targeted by glycogen synthase kinase-3 (T58), MAP kinase (S62) and Rho-dependent kinase (T71), respectively. The domains of Myc that interact with specific binding proteins are shown above the full-length protein structure. If the interaction results in Myc-dependent transactivation, the domain is represented in green. If the interaction results in Myc-dependent repression, the domain is shown as a red bar, and if protein interaction results in the repression of Myc function, the domain is represented in blue. Interactions that mediate both transcriptional activation and repression by Myc are indicated by a dashed bar. Domains of Myc that bind with partner proteins for which a role has not yet been determined are shown in grey. FBW7 is not a transcriptional cofactor, but is part of an E3 ubiquitin ligase that regulates Myc protein stability. SKP2 functions as part of an E3 ubiquitin ligase and as a cofactor for Myc. p300 is a histone acetyltransferase and p400 is a histone exchange factor. TIP48 and TIP49 are hexameric ATPases that are part of chromatin remodelling complexes, whereas TIP60 is a histone acetyltransferase complex. TRAPP, an adaptor protein, is the core subunit of the TIP60 and GCN5 complexes. Med, Mediator. (From Adhikary et al., 2005 with permission from Nature Reviews Molecular Cell Biology)

1.5 Myc control of cellular behaviors

Myc has an important role in regulation of different cellular activities, which can determine cell fate. Therefore, in normal cells Myc needs to be tightly regulated through Myc gene transcription as well as its protein stability and deregulation of its partner proteins[6, 31].

1.5.1 Cell cycle

Myc plays an important role in cell cycle regulation. Low Myc activity is associated with extended G1/G2 phases and delays of cells entering the S or M phases, respectively [38]. In

particular, Myc regulates cell cycle in the late G1 phase by inducing expression of the *Cyclin D1/D2/E*, *CDK 1/2/4* and *EF2* genes which are necessary for cells to enter S phase [39].

1.5.2 Apoptosis

In normal cells, inappropriate deregulation of Myc induces cell proliferation and growth but also apoptosis. Myc can induce apoptosis via both the intrinsic (p53-dependent) and extrinsic (mitochondrial) pathways [40]. Myc's effect on the intrinsic pathway involves either direct regulation of p53 or indirect regulation of p53 through P19^{ARF} up regulation[41, 42]. While p53 regulates pro-apoptotic proteins such as PIG3, PUMA and BAX it can also arrest the cell cycle by up regulation of the *P21CIP1* gene[43, 44]. In the case of the mitochondrial apoptosis pathway, Myc over-expression induces down regulation of BCL family proteins such as BCL2 and BCLXL that eventually leads to mitochondrial release of cytochrome c into the cytoplasm, thereby killing the cell[11].

1.5.3 Transformation

Myc overexpression alone may not be sufficient for cellular transformation, since Myc induced apoptosis plays a crucial role in preventing transformation [40]. Studies in lymphoma free individuals and Eμ-MYC transgenic mice suggest that factors in addition to translocation and over expression of Myc are needed [45-51]. Studies in rodent cells suggest that over-expression of either the RAS oncogene or anti-apoptotic protein Bcl2 can inhibit Myc induced apoptosis, leading to cellular transformation in cells with deregulated Myc [52, 53]. However, in transgenic mice models with tumors, Myc repression results in cell cycle arrest, differentiating back to normal appearing cells and apoptosis, which lead to tumor regression and extended life span [54].

1.5.4 Degradation

Myc transcriptional activation is finely tuned in normal growing cells. Myc over-expression may lead to Myc auto-regulation in a negative feedback loop, coupled with Myc ubiquitination and acetylation[6]. The Myc half-life time is short (approximately 24 min) [32], as it is a target for several E3 ubiquitin ligases such as Fbw7 and Skp2 [31]. Myc regulation by SCF^{Fbw7} ((SKP-Cullin-F-box) RING-FINGER domain ubiquitin ligase complex) [55] depends on phosphorylation of T58 and S62 amino acids in the MBI sequence, which is recognized by Fbw7[56, 57]. During growth factor stimulation of cells, S62 phosphorylation (pS62-Myc) increases Myc stability and activity [31, 58, 59]. However, phosphorylation of S62 facilitates T58 phosphorylation by GSK-3β [60]. T58 phosphorylation destabilizes Myc via Pin1 facilitated protein phosphatase 2A-B56a mediated dephosphorylation of S62 [61]. Singly phosphorylated Myc (pT58-Myc) is a substrate for Fbw7 ubiquitin ligase, which targets degradation by proteasomes (Figure 2) [30, 57]. Skp2 recognizes Myc in both N-terminal (MBII) and C-Terminal (HLH-LZ) domains and mediates Myc degradation independently of Myc phosphorylation status (Figure 2)[62, 63]. On the

other hand, Skp2 can promote Myc transcriptional activity and induce S phase entry of the cells [63], which suggests a constitutive role for Skp2 promoting Myc activities and controlling Myc level at the same time. Myc interacts with protein acetyltransferases such as CBP/p300 and Tip60[64, 65]. Myc acetylation in multiple residues can increase its stability and interfere with its ubiquitin-mediated degradation[31, 64, 66]. Moreover, Myc expression can directly regulate the *SIRT1* gene, SIRT1 is a protein deacetylase, which can de-acetylate Myc (unstable form) and thus controls its level in a negative-feedback loop [67].

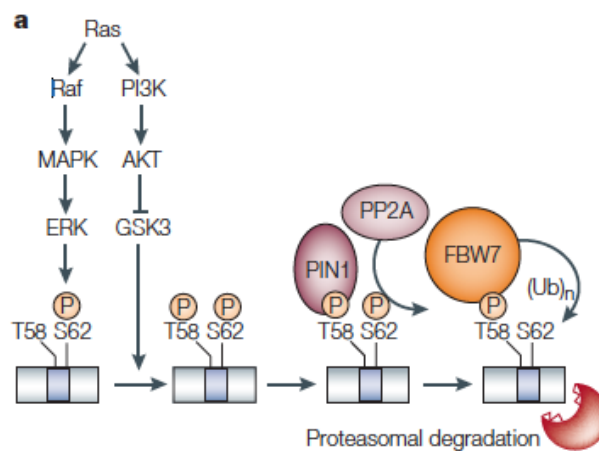


Figure 2. Regulation of Myc function by Ras dependent signalling pathways. The stability of Myc is regulated by the phosphorylation of residues Ser62 (S62) and Thr58 (T58) following stimulation by Ras-dependent signalling cascades. Ras stabilizes Myc by phosphorylation of S62 through the MAPK/ERK pathway and by the inhibition of glycogen synthase kinase-3 (GSK3) through phosphatidylinositol 3-kinase (PI3K) signalling. GSK3 can phosphorylate T58 if S62 is phosphorylated. Phosphorylated T58 is recognized by the prolyl isomerase (PIN1), which leads to the isomerization of Pro59 and enables phosphatase-2A (PP2A) to remove the phosphate residue from S62. SCFFWB7 (FBW7 in the figure), a ubiquitin E3 ligase, polyubiquitylates Myc in a T58-phosphorylation dependent manner and labels it for degradation by the proteasome (From Adhikary et al., 2005 with permission from Nature Reviews Molecular Cell Biology)

2 MYC EVOLUTION, SEQUENCE OR PROTEIN STRUCTURE?

2.1 Myc evolution and protein sequence

Transcription factors are important coordinators of development in multicellular organisms, and are involved in regulation of their biological processes. Transcription factors exhibit a complex network of interactions with proteins and DNA. Subsequently, TFs are well known to recruit other co-factors as well as themselves being recruited by their partner proteins to specific sequences of target genes, thus determining patterns of gene expression.

Understanding a transcription factor family and its conserved regulatory motifs in different species, can give key understanding of its role in regulating mammalian cell fate and function as well as the altered pathways that characterize cancer cells.

Myc belongs to a super family of bHLH proteins, which can bind to heterodimerization partners via their bHLH domain and play important roles in development of both vertebrate and invertebrate metazoan organisms [68, 69]. Recent genome sequencing and functional studies in the unicellular *Monosiga brevicollis* report a functional Myc homolog and expand these ancient TFs to a time before the evolution of metazoans [70, 71]. Previous studies report the C-terminal bHLH-LZ domain (CTD) as the main evolutionary conserved region of all Myc family proteins [72]. Moreover, vertebrates have a distinctive conserved site in bHLH DNA binding and dimerization region [73].

Myc family proteins are highly diverged TFs in the N-terminal sequences that lack high propensity for structure formation. Due to this fact, using classical protein multiple alignment methods gives a poorly connected phylogenetic relationship for Myc family proteins outside mammalian organisms. However, previous phylogenetic studies of bHLH TFs used C-Terminal conserved domains of Max family proteins [72]. These are more successful at grouping Myc proteins in relation to more distantly related bHLH TFs but they may of course lack important information needed for accurate phylogenetic characterization within the Myc protein family. Myc proteins have 6 blocks of conserved residues through their protein sequences [13]. In vertebrates, Myc proteins have 3 main highly conserved blocks in the N-terminal region, which are also found in most Eumetazoans except within the Mandibulata (e.g. fruit fly species). However, fruit fly dMyc has been experimentally shown as a valid Myc homologue. The *dmyc* gene is important for regulation of growth and development, and dMyc deregulation is correlated to changes in both fruit fly cell size as well as their body size [74]. Outside vertebrates, one or more motif matches are insufficient for reliable detection of Myc proteins. Furthermore, the N-terminus of choanoflagellate (*M. brevicollis*) Myc proteins is generally very poorly connected to mammalian Myc proteins but interestingly it does have some similarity to MBI and MBII residues (Paper I), which provides some evidence that

these regions of Myc proteins may have a conserved function throughout the evolution of Myc proteins [70-72].

2.2 Intrinsic disorder profile of proteins, new insights on evolution

Myc as well as other transcription factors (TFs), has a certain region within their protein sequences or the whole TF, lacking defined structural conformation in physiological conditions[75]. In general, proteins can have well-structured domains as well as unstructured regions that are termed, Intrinsically Disordered Region (IDR) [76]. Proteome-wide analysis shows that many of the TF interaction domains are within IDR regions of the proteins[77]. Consistently, previous reports imply that, upon binding to other partner proteins IDRs could locally fold, such that appropriate conformation changes can induce protein-protein complex formation[76]. More recently, short structural elements (10-20 amino acids), which are functional in protein-protein and protein-DNA recognition of the proteins, have been identified in proteins [78, 79]. These elements, MoRF (Molecular Recognition Feature)/ANCHOR, can be conformationally stabilized by partner proteins and thus transit from disordered to an ordered conformation [77, 79]. Thus, this flexible nature of IDR regions is important for high specificity and low affinity of TF binding to multiple bind protein partners[77].

Furthermore, recent studies identified an evolutionary conserved pattern for IDR regions in mammals, suggesting the core functional definition of a protein is conserved through its rapid IDR evolution [80].

3 RIBOSOME BIOGENESIS

Ribosomal RNA (rRNA) is the major product of RNA synthesis in growing mammalian cells. In actively growing cells, about 35% of transcriptional activity is devoted to rRNA synthesis and a large part of the mRNA transcription pool encodes proteins essential for rDNA transcription and ribosome assembly [81, 82]. Together, these transcripts unite to provide the machinery needed for protein synthesis and cell growth [82].

In mammals, ribosomes have two distinct subunits referred to 60S and 40S. 60S is the larger subunit of the ribosome and contains around 49 ribosomal proteins together with the 28S, 5.8S and 5S rRNAs [83, 84]. On the other hand, the smaller 40S subunit consists of 33 ribosomal proteins and the 18S rRNA [83, 84]. rDNA is transcribed by RNA Polymerase I (Pol I) and Pol III. Whereas, RNA Polymerase III transcribes 5S rRNA, Pol I transcribes a 47S precursor ribosomal RNA which is later cleaved to the mature 18S, 5.8S and 28S rRNAs through exo- and endo-nucleolytic cleavages [85]. RNA Pol II transcribes ribosomal protein genes and other additional factors [86].

The nucleus contains multiple copies of rRNA genes (Figure 3). The genomic regions containing rRNA genes in human cells are clustered in the nucleolar organization region (NORs) of chromosomes 13, 14, 15, 21 and 22, which combine together in growing cells to constitute the nucleolus [87-90]. Importantly, morphological and functional changes in the nucleolus are diagnostic markers of particular tumor types and states [91]. These morphological changes include the number of nucleoli per nucleus, nucleolus size and shape as well as its “chromatin texture” [91]. Changes in nucleolar phenotype are the consequence of an increasing need of ribosome biogenesis in highly proliferative cells including cancer cells, and changes in rRNA expression appear to be an important mechanism in tumor genesis [91].

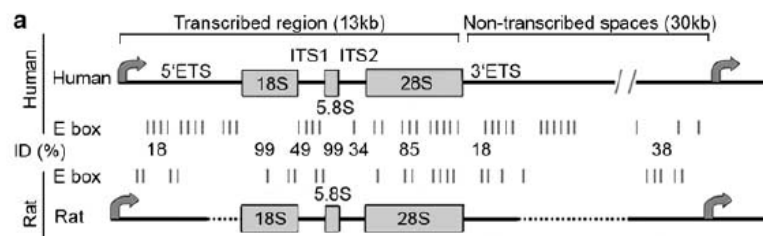


Figure 3. c-Myc binds to rat rDNA and directly induces rDNA expression (a) The sequence and distribution of E-boxes in the rDNA is strongly divergent between humans and rats. The upper diagram shows human rDNA derived from GeneBank no. U13369, containing the transcribed and untranscribed regions of the rDNA repeat unit. The lower diagram shows rat rDNA, derived from accession numbers X04084, X03838, X61110, X00677, X16321, V01270 and X03695. Dotted sections of the diagram represent regions where rat sequence information is incomplete. Locations of E-boxes are marked with vertical lines. The percentage identity between human and rat sequences is noted for each coding and non-

coding section. Horizontal bars represent regions amplified in chromatin immunoprecipitation (ChIP) assays in rat cells, and labels indicate their approximate position relative to the transcription start site (R0). (Shiue *et al.*, 2009 with the Permission from *Oncogene* “springer Nature”)

3.1 Mammalian nucleolus structure and rDNA

Mammalian cells have three main nucleolus compartments including Fibrillar Centers (FCs), the Dense Fibrillar Component (DFC) and the Granular Component (GC) (Figure 4) [86, 92]. Transcriptionally active cells have different numbers of small FCs, while cells with reduced protein biogenesis have one large FC [93]. Nascent rRNA transcription occurs in the region between FCs and DFCs, causing the later accumulation of pre-rRNA (47S) in the DFC compartment where nucleolytic cleavage of the 47S transcript occurs (Figure 4) [94]. Moreover, intra-nucleolus migration of the 18S, 5.8S and 28S rRNA and synthesis of ribosomal subunits occurs in the GC [94].

The diploid human genome contains around 400-repeated ribosomal RNA genes, mainly arranged as tandem head to tail clusters, from telomere to centromere, in NORs [85, 90, 95]. In human cells, each 43 kb rDNA repeat has an approximately 13kb sequence encoding pre ribosomal RNA (47S) and each repeat is separated from the next by ≈ 30 kb of intergenic spacer (IGSs) DNA sequence (Figure 3) [90]. However, both transcribed and non-transcribed sequences of pre-rRNA are necessary for recognition and binding of the Pol I pre-initiation complex and subsequent rDNA transcription [81]. In particular, the regions including proximal-spacer promoters, enhancer and terminators have been shown to be important [96].

3.2 Nucleolar scaffold and DNA compartmentalization

Early genome organization and electron microscopy studies suggest a dynamic link between nuclear structures and gene expression [97, 98]. The nucleus has structural compartments including nucleoli, the nuclear envelope and the nuclear matrix. Further, the organism's chromosomes are located within the nucleus. Association of a genomic location with nuclear structures can influence the transcription of that genomic region [99]. Disorganization of nuclear structural compartments, including the nucleoli, nuclear envelope and nuclear matrix is associated with cancers [100, 101].

Within the nucleus, genes inside widely separated chromosome territories can loop and co-localize into structurally distinct active transcription sites [102, 103]. In these active transcription sites, often termed “transcription factories”, DNA loops are surrounded by transcription factors, RNA polymerases and other auxiliary co-factors [104].

Nuclear matrix is an intra-nuclear structure that provides a protein-supporting scaffold for the diverse processes and functions, including gene looping, DNA transcription and RNA processing [97, 105]. DNA sequences, which mediate loop formation, are known as Scaffold Attachment Regions or Matrix Association Regions (S/MARs) [105]. Accordingly, multiple

loop structures are maintained as domains due to end attachment of the S/MAR regions to the nuclear matrix scaffold [106-110].

Interestingly it has been speculated that repetitive non-transcribed elements of rDNA genes (IGS segments) can play a role in the induction of higher-order rDNA structure [85], which will be discussed in section 3.4 . Moreover, it has been shown by three-dimensional electron microscopy that compact active rRNA genes within fibrillar center (FCs) have loop structures in the form of the well-known Christmas tree structure [111].

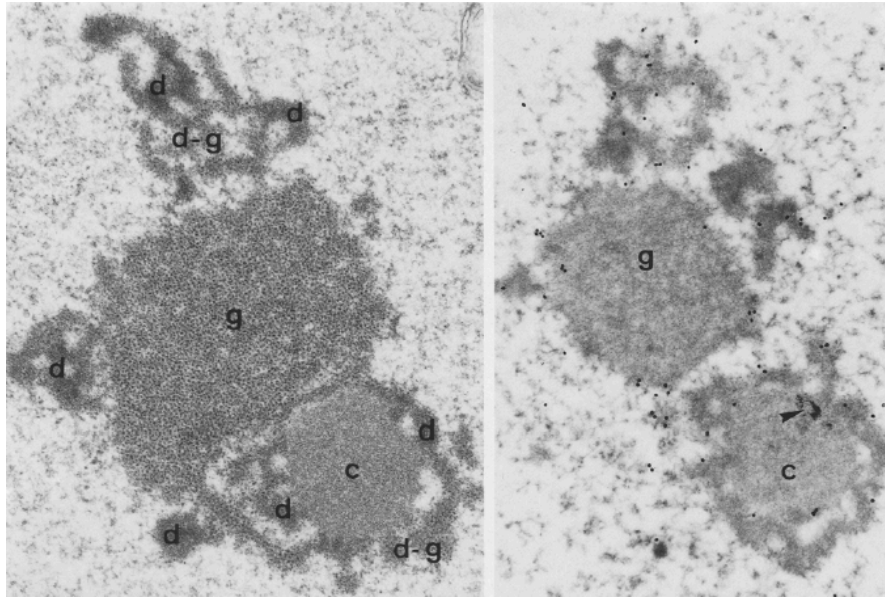


Figure 4. **LEFT:** Nucleolus of a human Sertoli cell. Technique as in Fig. 2 with additional acetylation. The single large fibrillar centre (c) is surrounded by strands of dense fibrillar component (d) sometimes intermingled with granular component (d-g) Dense fibrillar component is also found at the periphery of a distinctly defined field of granular component (g) and at a great distance from the fibrillar centre, x 30,000. **RIGHT:** Human Sertoli cell nucleolus. Electron microscopic in situ hybridization of rDNA. The 5-nm gold particles of the label are marked with Indian ink. The original preparation is published in Schwarzscher and Wachtler (1991). Density of label above background is seen only in the region of the dense fibrillar component. x 30,000. (Hans Georg Schwarzscher and Franz Wachtler with (1993) with permission from *Anatomy and Embryology*)

3.3 Ribosomal RNA gene transcription

The rDNA promoter consists of two main binding elements, the upstream control element (UCE) and the core promoter [112]. In general, transcription initiation of rRNA genes needs assembly of specific complexes at their promoters including, RNA Polymerase I, upstream binding factor (UBF), transcription initiation factor 1A (TIF1A), and the selectivity complex 1 (SL1) [81]. The most important step in rRNA gene transcription is the recruitment of Pol I to the rDNA promoters, which is a limiting factor in pre-ribosomal RNA (47S) biogenesis [81]. Pol I recruitment to the rDNA promoters is achieved by its recruitment to the Pol I pre-initiation complex[81]. Such complexes are formed by UBF and SL1 interacting together, whereas UBF alone can bind to the multiple sites on rDNA repeats, via the minor groove of the DNA, and forms a nucleosome like structure and can recruit the other complexes [81,

113]. Upon UBF-SL1 complex binding to the UCE and the core promoter of rRNA genes, it can further interact with and activate TIF1A [95, 114, 115]. Thereafter, the pre-initiation complex can recruit Pol I to the rDNA promoters and initiate 47S ribosomal DNA transcription [81]. The 47S precursor ribosomal RNA is later cleaved and generates the mature rRNA subunits (18S, 5.8S and 28S) [85].

RNA Polymerase III transcribes the fourth rRNA subunit (5S rRNA) [81, 116]. Pol III transcribed genes require (Transcription Factor for Polymerase III) TFIIC-TFIIB recruitment to their promoters [117]. However, TFIIC- has a low affinity for 5S rRNA promoters and Pol III transcription is therefore dependent on TFIIIA binding to the 5S RNA promoter regions [81]. Ordered recruitment of TFIIIA, TFIIC and TFIIB to the transcription start site are required for Pol III loading and transcription initiation of 5S RNA [81, 117].

3.4 Role of Myc in rDNA transcription

In eukaryotes, ribosome biogenesis requires all three major RNA polymerase activities. Pol I, only transcribes rRNA genes and produces pre-RNA (47S). Pol III, transcribes some small non-coding RNAs as well as 5S rRNA and transfer RNAs (tRNA), both of which are essential for ribosome biogenesis and protein production [81]. Pol II, transcribes genes (mRNAs) that encode ribosomal protein subunits, accessory factors, Pol I and Pol III subunits and their co-factors [81]. Pol I and Pol III transcriptional activates represent the majority of nuclear transcription [81]. Moreover, transcription of Pol I and III subunits and co-factors increase in different cancer cells [81, 116, 118]. On the other hand, Myc oncoprotein and other tumor suppressors can modulate rDNA transcription [116, 118]. Accordingly, *Drosophila* Myc (dMyc) was found to play a role in ribosome biogenesis [119]. Moreover, a recent microarray study, elucidated ribosome biogenesis as the Myc core signature in cancer cells, independent of cell type and species [120]. Myc's role in ribosome biogenesis, by regulating transcriptional activity of all three RNA polymerase is briefly described in Figure 5.

We, at the same time as another group, found that Myc regulates Pol I, independently of its Pol II transcriptional activities [20, 21, 120]. Moreover, we revealed that Myc could localize to the nucleolus, bind to E-box motifs and promote rDNA transcription [20]. In Myc-ER rat fibroblasts, where Myc is fused to part of the estrogen receptor and is dependent on addition of estrogen agonists for its activity, absence of Myc activity reduces rDNA transcription by 15-20% compared to the wild type cell line (TGR1) [38]. Addition of agonist ligand to Myc-ER cells to induce Myc activity increased rDNA transcription [121]. Myc can regulate pre-rRNA synthesis by direct binding and regulation of UBF, SL1 and Rn3 (TIF1A) complexes on the rRNA genes [81]. However, a variety of Chromatin immune precipitation (ChIP) based studies show Myc binding to the UBF and Rn3 (TIF1A) gene promoters [122, 123]. In the mouse model, Myc expression induced Pol I specific subunit transcription as well as transcription of genes encoding the UBF and SL1 subunits (TBP, TAF1C, TAF1D) [122]. In

addition, Myc can recruit TRRAP (TTransformation/tRanscription Associated Protein), which has a histone acetyltransferase co-activator function, via the MBII domain [22, 124]. Interestingly, Myc localization to the nucleolus can influence the Pol I target gene by altering its histone H3 and H4 acetylation, in both the transcribed and non-transcribed (IGS) regions of the rRNAs genes [20, 21].

As for Pol I, Myc can regulate around 40% of Pol III subunits as well as its target genes including 5S rRNA [81, 116, 122]. The transcriptional level of 5S rRNAs increases upon Myc induction in different cell types and tumors [125-127]. Myc can directly influence Pol III transcription by its interaction with the TFIIIB subunit known as Brf1, which is the limiting factor of Pol III transcription [125, 128]. Importantly, Myc increases histone H3 and not H4 acetylation on pol III target genes, possibly via the Gcn5 acetyltransferase that associates with the TRRAP co-factor [124].

As previously discussed, rRNA genes within fibrillar center (FCs) can form loop structures [111]. And recently, we have discovered that Myc could play an alternative role in ribosomal biogenesis by inducing loop structures and thereby altering the higher order chromatin structure of the rDNA repeats of growth stimulated cells [121]. Upon serum stimulation of quiescent rat fibroblast cell lines, which induces Myc level and leads to a subsequent increase in the binding of the Pol I and TBP (an SL1 complex subunit) in both upstream core promoter sequences and downstream of the transcribed region in the IGS region [121]. Thus, c-Myc can induce a chromatin hub, where rRNA gene loops are linked by promoter and terminator regions in close proximity that could be a mechanism for enhancing re-initiation of rDNA transcription.

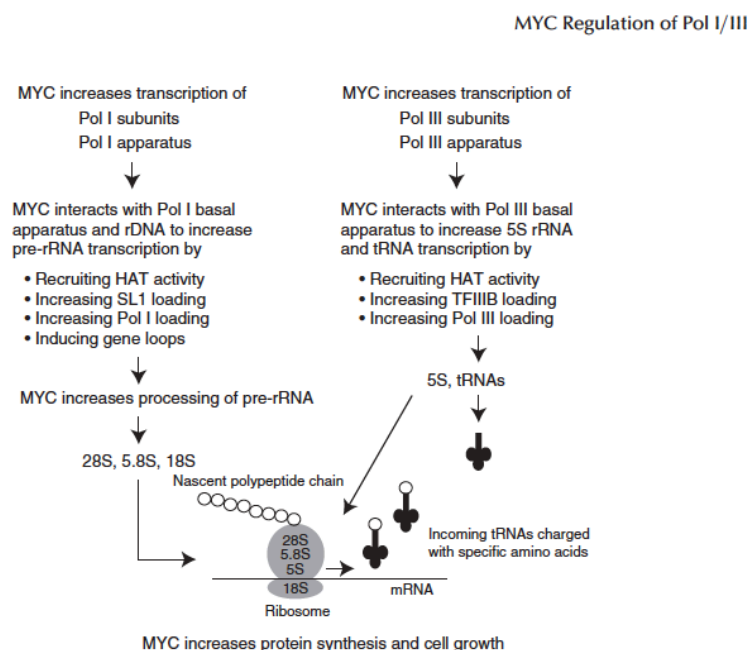


Figure 5. Myc regulates Pol I and Pol III at multiple levels to stimulate protein synthesis and cell growth (Kirsteen J. Campbell and Robert J. White (2014) with copyright permission from Cold Spring Harbor Laboratory Press Cold Spring journal).

4 ROLE OF MYC IN LYMPHOMA MALIGNANCIES

4.1 Lymphoma malignancies

Lymphocytes are fundamental immune cells circulating in blood and the lymphatic system. In health (non cancer) individual B (bursa or bone marrow) and T (thymus) lymphocytes represent 20 to 45% of white blood cells [129]. B and T cells are an important component of the human innate and adaptive immune systems [130, 131]. Lymphocytes are the origin of a diverse group of hematological malignancies, which are classified by their cells of origin and their pathological characteristics [132].

According to the World Health Organization (WHO) in 2016, lymphoid neoplasms are classified into five major groups. These are 1. Mature B-cell lymphoma neoplasms including Burkitt's lymphoma, 2. Mature T and NK cell neoplasms, 3. Hodgkin-lymphoma, 4. Post transplants lymph proliferative disorders (PTLD) and 5. Histiocytic and dendritic cell neoplasms [133].

4.1.1 T cell lymphocyte

Bone marrow is the niche where most hematopoietic cells are generated and complete their development. Unlike other bone marrow (BM) derived hematopoietic cells, T-cell progenitors migrate into the thymus and further undergo lineage commitment, selection and maturation [134, 135]. Several developmental stages of T lymphocytes are phenotypically characterized by their patterns of co-receptor expression. T-lineage progenitors give rise to the CD4⁻CD8⁻ double negative thymocytes, in which T-cell receptor recombination starts and T-cells start the maturation stage and become double positive (DP) CD4⁺CD8⁺ thymocytes [134, 135]. At this stage, DP cells have rearranged TCR α genes and undergo CD4 or CD8 lineage transformation that includes both positive and negative selection processes [134, 135]. Such mature T-cells (Naïve T-cells) are then released into peripheral blood [135].

In peripheral blood T-cells account for around 75% of the lymphocytes and T-cell activities play a major role in cell-mediated immune responses [129]. Naïve T-cells are mainly classified into three different lineages including; 1. CD4⁺ T-helper cells, which are mainly involved in immune responses by releasing cytokines 2. CD8⁺ T-cytotoxic cells, which directly destroy their targets by antigen recognition via their surface receptors directed against cancer cells or infectious agents 3. T-suppressor cells, a regulatory T-cell lineage that is important in autoimmune tolerance against self antigens [130, 136, 137].

4.1.2 B-cell lymphocytes

The B-cell lineage is derived from the development of hematopoietic stem cells in the bone marrow niche, and plays a major role in adaptive immunity [131, 138]. The B-cell system is considered as a positive regulator of immune response by antibody production [129, 131].

Each B-cell can proliferate in response to specific stimulation and later produce plasma cells, which are able to produce only one species/type of antibody [131, 139]. However, B-cells can function in different immunological responses including antigen presentation activities in the lymph node [140] as well as in cytokine production (e.g. IL-10, TNF α , IL-6 and IFN γ) [141, 142].

In human, immature B cells pass different stages of developments in BM independently of antigen stimulation [131, 139]. In particular, progenitor B-cells (pro-B-Cells) mainly require hematopoietic stem cell stimulation by stromal cell products, mainly interleukin 7 (IL-7) and activation by fms-like tyrosine kinase 3 [139]. During BM development stages, pro-B-cells differentiate to precursor B-cells, in which B-cells acquire their sequential light and heavy chain rearrangements, leading to formation of the B-cell Receptor (IgM). At this point they have reached the immature B cell stage [131, 138, 139]. At the immature stage, B-cells leave the BM and enter the periphery blood, where they further complete their development to mature B cells co-expressing IgD and IgM on their cell surface [131, 139]. Mature B cells migrate to the lymphoid organs, where as their differentiation to mature memory B cells or antibody secreting plasma cells occurs, in antigen dependent manner [138, 139].

4.2 Introduction to Burkitt's Lymphoma

In 1985, Dr Denis Burkitt first characterized a tumor in children, which was later identified as a pathological syndrome, known as Burkitt's Lymphoma (BL) [143]. BL is a highly proliferative non-Hodgkin's B-cell lymphoma, which histologically has multiple nucleoli per nucleus [144]. BL has three epidemiologically recognized variants, including endemic BL, sporadic BL and a HIV/EBV associated immunodeficiency form [145].

In endemic areas BL accounts for around 50% of children with non-Hodgkin lymphomas and is mostly located in facial bones [144]. Moreover, the endemic variants are associated with either or both EBV and malaria infections [146]. In comparison, sporadic BL represents around two percent of adult lymphomas in develop countries and they mostly involve lymphoid tissues of different organs including the gut and the respiratory tract [146].

4.2.1 Myc regulation in Burkitt's Lymphoma

Early chromosome karyotyping observations in BL tumors revealed an additional band at the end of chromosome fourteen, as well as the absence of a band in chromosome eight [147]. This t(8;14)(q24;q32) translocation is observed in more than 85% of BLs, independently of their origin [146]. Later, independent groups found that the region containing the *myc* gene (8q24) was translocated to the immunoglobulin heavy chain locus [2, 148] or to the immunoglobulin light chain locus [149].

Myc deregulation resulting from the t(8;14)(q24;q32) translocation is the diagnostic hallmark of BL tumors. *myc* gene translocation and deregulation sometimes followed by its

rearrangement or mutation is associated with aggressive lymphoid neoplasms, including BL but also diffuse large B-cell lymphoma (DLBCL) [147, 150]. However, Myc translocation per se seems to be insufficient for tumor genesis in lymphoma-free individuals and cells need to overcome other inhibitory effects before they can develop into lymphoma cells [46]. In particular, deregulation of BCL6 and BCL2 are found in aggressive B cell malignancies, known as Double Hit Lymphoma (DHL). Moreover, Myc and p53 mutations are identified in more aggressive BL, and p53 mutation in DHL, harboring deregulated Myc and BCL proteins is associated with even more aggressive lymphoma [151, 152].

In addition to the chromosomal abnormalities, 60% of BLs have missense mutations in Myc clustered in functional domains, primarily Myc Box I (MBI) [32, 147]. This suggests a selective adaptation of the Myc protein in Myc driven lymphomas. MBI Mutations, including E39D, A44V, P57S, T58I, T58A, can affect the conserved MBI region and alter the predicted conformation of the Myc protein in some cases (Paper I) [35, 153].

Previous studies in rat fibroblast revealed that these mutations have different oncogenic effects on the Myc protein[35]. Importantly, Threonine 58 to Alanine substitution (T58A) enhances Myc transformation activity while it reduces Myc induced apoptosis [35]. In contrast, T58 to Isoleucine substitution (T58I) has the opposite biological effect on Myc, by reducing its transformational activity and having no noticeable difference on Myc-dependent induction of apoptosis, compared to a cell line expressing wild type Myc [35]. In an apparent contradiction to this, T58I is the most reported mutation in BLs, suggesting that different substitutions can be a selective step for Myc in different lymphoma cells.

5 MATERIAL AND METHODS

5.1 Lentiviruse vectors constructions

Wild type c-Myc, T58AMyc and T58IMyc DNA was generously provided by M.D Cole (Department of Genetics, Geisel School of Medicine at Dartmouth, USA) cloned in Cβh-Myc vectors (Figure 6) that were first sent for sequencing using a 5'ATGCCCCTCAACGTTAGCTTC 3' primer to confirm the correct wild type or mutant sequences. The human WTMyc, T58A and T58I sequences were amplified by PCR using SK primer sets; SK-FW 5' TCTAGAACTAGTGGATC 3' and SK-Rev 5' CGAGGTCGACGGTATCG 3' containing BamHI digestion sites.

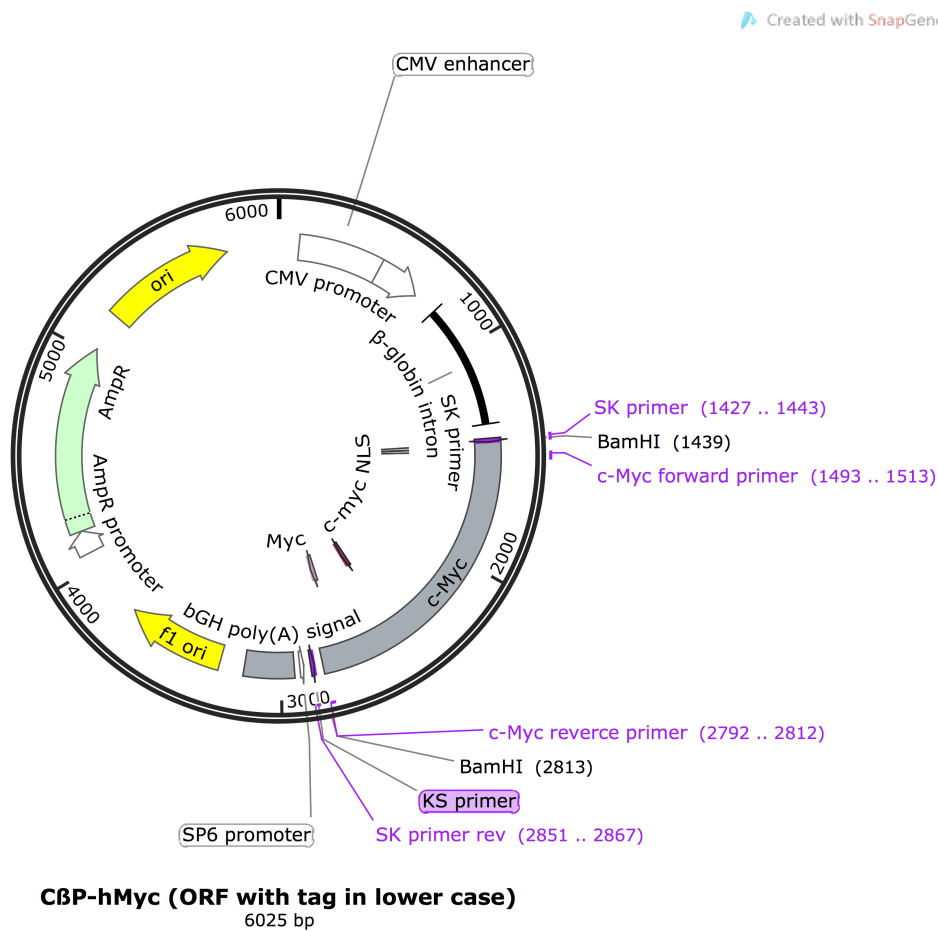


Figure 6. Illustrating Cβ-hMyc plasmids containing WTMyc or Mutation including : T58AMyc and T58IMyc

Accordingly, we opened pSIR-TREMYC-IRES-EGFP-PGK1-rtTA2 vector (generously provided by Kari Högstrand) using *BamHI* enzyme digestion at a unique cloning site. Amplified human WTMyc, T58A and T58I sequences were then cleaved with *BamHI* and cloned into *BamHI* cleaved pSIR-TRE-IRES-EGFP-PGK1-rtTA2 lenti-virus expression

vector, generating pSIR-TRE-MYC-IRES-EGFP-PGK1-rtTA2 vectors, as illustrated in Figure 7.

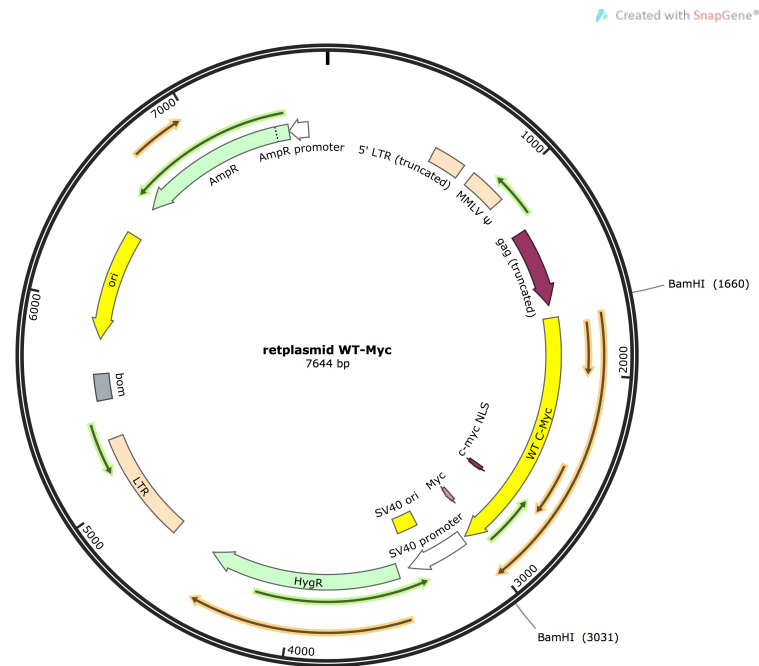


Figure 7. pSIR-TRE-MYC-IRES-EGFP-PGK1-rtTA2 plasmid containing WTMyc , T58A or T58I fragments and other respective sequence regions , used for retroviral transductions

5.2 Making the triple hit Myc-driven B-cell lymphoma-like cell system.

Establishing a cell line with patient lymphoma cells is often difficult, since cells might not yet have acquired all genetic alterations require for independent cellular growth *in vitro* [40]. Moreover, inducing Myc in primary B-cells results in enhanced apoptosis. However, experimental approaches have been found by our collaborator (Alf Grandien), by using anti apoptotic proteins BCLXL and BMI1, which when overexpressed together with Myc allow propagation of primary B-cells, *in-vitro*. Briefly, The Lipofectamine 2000 system (Life Technologies, Paisley, UK) was used for transient transduction of the Phoenix-Eco packaging cells (kindly provided by GP. Nolan, Stanford University). Retroviral particles were obtained and concentrated by centrifugation (6000 x g, overnight at +4°C). LPS stimulated B-cells were transduced by retroviral pool spin infection in 8ug/ul of polybern (Sigma-Aldrich). Each retroviral pool contained viruses containing one of WTMyc, T58AMyc or T58IMyc together with both BCXL and BMI1. Cells were re-plated in complete culture medium supplemented with 2mg/ml Doxycycline. B-Cell Lymphoma-like development and survival depends on retroviral transfection with all of these three genes that alter the cell fates by overexpressing Myc and inhibitors of the intrinsic and P53 apoptosis pathways.

5.3 Design pipeline for read alignments and subtraction of human reads from mouse reads

We used human c-Myc, BCLXL and BMI1 for transduction of the mouse splenic B-cells. Subsequently, to omit the influence of these 3 human overexpressed genes, we made a human mini-reference genome that was used to subtract reads of human origin. Read quality was assessed by Fastqc (v.0.10.1) and the reads were then aligned to the mouse genome, (build GRCm38.75, obtained from the Illumina iGenomes with ensemble annotation files and pre-built Bowtie2 indexing) using the splice-aware short read aligner Tophat2 (v.2.0.4) with Bowtie2 (v.2.2.6), and default options with $-G$ indicating the GRCm38.75 GTF file and $-g$ set to 1 [154, 155]. Next, the raw reads for each sample were aligned by Tophat2 to a mini-reference genome and the aligned reads were subsequently subtracted from the bam files containing reads aligning to GRCm38.75.

5.4 Using PONDR®VLXT for prediction of BL-AM protein disorder profiles

PONDR®VLXT [78, 156] predicts the disorder profile of c-Myc protein and gives residue-by-residue predicted disorder scores. However, the PONDER-VLXT algorithm can also predict short ordered regions within an intrinsically disordered region. This additional feature of this algorithm makes it more suitable for analyzing the Myc associated cancer mutations. Briefly we substituted residue-by-residue disordered values into the equivalent position of each amino acid and data were plotted using ggplot2 package in R [157].

5.5 Analysis of the nuclear matrix and DNA methylation assay

Matrix attachment region or scaffold attachment regions are AT-rich sequences, which define DNA sequences that bind to the extracted nuclear matrix, and are generally located in regulatory elements of the DNA [158, 159]. There are different methods for MAR isolation. In general, DNase I or a combination of restriction enzymes is used to digest nuclei. Next, histones and other proteins are extracted by high salt concentration (2M NaCL) or lithium diiodosalicylic (LIS). Thereafter, DNA sequences can be separated into an insoluble portion, containing matrix associated DNAs, and a soluble portion containing non-matrix associated DNA. Importantly, some methods suggest DNase I treatment of the sonicated nuclei before sucrose gradient centrifugation, in order to avoid loss of nucleoli due to their being trapped within a chromatin network [160]. Also, DTT addition to both digestion and histone extraction steps is reported to promote the solubilization of the matrix and prevent non-specific DNA binding to the nuclear matrix[161].

In Rat Fibroblasts, we measured the methylation level of matrix-associated regions in the CpG site at the -143 position upstream of the transcription initiation site. Matrix associated and non-matrix associated DNAs were recovered after digestion with *BamHI* and *XhoI*, using additional chloroform washing steps, in order to remove residual phenol that was inhibitory to the methylation sensitive enzyme digestion. *HpaII* and *MspI* have the same recognition site

(5'...CCGG...3'), but the *HpaII* enzyme cannot cut methylated DNA. Therefore, we used this enzyme sensitivity assay and evaluated the level *HpaII* resistant sites by qPCR using the forward primer 5' AGCATGGACTTCTGAGGCCGAG 3' and the reverse primer 5' CATAAAGCTGCCCCAGAGAG 3'.

6 AIMS

- To determine the extent of conservation/ divergence of intrinsic protein disorder throughout the Myc protein family and to evaluate the effects of cancer-associated substitution mutations on the intrinsic disorder profile of.
- To determine whether and how Myc changes the higher order chromatin structure of the rDNA by modulating its interaction with the nuclear matrix.
- To identify gene targets of Myc that are important during development of a lymphoma phenotype in B-cells as well as how this is augmented by cancer-associated Myc mutants.

7 RESULTS

7.1 Paper I: phylogenetic and molecular evolution studies of the Myc family proteins.

To understand Myc evolution and its role in cancer, we used the intrinsic disorder profile of Myc family proteins as a signature to analyze their structural evolution, as well as the adaptive effect of Burkitt Lymphoma-Associated Mutations (BL-AM) on the intrinsic disorder profile of the c-Myc protein. In particular, we revealed that the IDR profiles of Myc proteins are a primordial signature within the evolution of Myc in organisms ranging from the Protista to the broad range of Metazoans. Further, our findings suggest that adaptive BL-AM can affect the intrinsic disorder profile of the c-Myc protein.

To understand Myc structural evolution, and the functional role of its adaptive cancer mutations we first used BLASTP and identified representative metazoan Myc protein sequences in the UniProt Ref50 dataset. Here we illustrate UniRef50 Myc group's similarities to several defined c-Myc (P01106) signatures. And we further run multiple-alignments for phylogenetic analysis of the UniRef50 groups (representative Myc protein sequence) in Metazoans, where our dendrogram trees show the clades for c-Myc, L-Myc, N-Myc and d-Myc (fruit fly-Myc). However, most other Myc proteins are poorly connected to the dendrogram with low statistical support.

To have a better understanding of Myc protein evolution, we used an alternative approach by comparing Myc proteins by their predicted intrinsic disorder profiles. Thus, we used a variety of algorithms for in-silicon prediction of the intrinsic disorder region profiles (IDR) of Myc proteins. We revealed that all Metazoan Myc proteins are highly disordered proteins. Subsequently, we used a systematic analysis of Myc disorder predictions (IDR) and the distance between the scores for different Groups of UniRef50 representative Myc proteins for phylogenetic analysis. Interestingly, using IDR scores inserted into multiple alignments of Myc proteins, we found a better dendrogram tree grouping most Myc proteins in individual clades with good support. As the results illustrate, there is good separation of c-Myc, N-Myc and L-Myc clades, similar to the sequence based tree, human c-Myc (P01106) and Xenopus Myc(Q05404) were both placed outside the N-Myc clade. Other Myc proteins, which are functionally related to human Myc proteins, grouped together well in clades and were well connected to the clades of established Myc proteins.

Myc Box I is a main functional domain of the Myc protein and is enriched with mutations found in BL. The known BL-associated mutations are located in the disordered regions of Myc and have different functional roles in cell proliferation as well as cellular transformation and apoptosis [32, 35]. One of the important parts of my thesis was first to analyze the effect

of BL-associated mutations on intrinsic disorder structure of the c-Myc protein and to further study their adaptive effect on Myc functionality in my last project (Paper III). We used the PONDR®VLXT [78, 156] algorithm to predict and visualize protein disorder profiles of wild type Myc and different reported BL mutations known to alter Myc function [35]. Our *in silico* finding suggests that adaptive amino acid substitutions can locally effect the disorder of the Myc protein, and that different amino acid substitutions of the same residue (Threonine 58) effect the Myc IDR profile differently. This finding was interesting in relation to their well-described functional differences in Rat fibroblast systems [32, 35]. Therefore, analyzing these missense mutations in a relevant B-cell system (Paper III) could help to couple differences in the effects of BL mutation on protein conformation to differences in their effects on protein function.

7.2 Paper II: functional analysis of Myc family proteins in mammalian cells.

In this study, we found a novel c-Myc role in ribosome biogenesis by its role in the enrolment of the rRNA genes to the nucleolar matrix. First, we predicted the potential Matrix Attachment Regions (MAR) in rDNA repeats, using the MAR-wiz computational tool. This exclusively identified the InterGenic Spacer (IGS) sequences, as the most prominent Matrix/Scaffold attachment sites for mammalian rDNAs. We then used Human (HeLa) and Rat (TGR1) cell lines, from which we isolated nucleoli by sucrose gradients, before further digesting the isolated nucleoli with DNase I and RNase A and PCR (Polymerase Chain Reaction) identification of the possible matrix protected rDNA regions. Using different primer sets across the 40kb rDNA sequence, we showed that the IGS regions have a high tendency for matrix association.

We further monitored matrix attachment in response to the growth stimulation by serum. The matrix-protected region was significantly reduced in starved cells compared to the serum-stimulated cells. However, the H16, H27, H37 and H40 regions were matrix protected in both serum starved and stimulated cells. Serum stimulation of the cells activates Myc. Therefore we inhibited Myc by treatment with a small molecule drug (10058-F4). Interestingly, Myc inhibition resulted in abrogated nuclear matrix protection in growth-stimulated cells, which suggests a role for Myc in matrix attachment of rDNA in actively growing cells.

To further characterize S/MAR sequences and their potential matrix binding abilities in the formation of chromosomal loops, we established a MAR-Loop method in our lab by combining the MAR assay with the 3C assay without the formaldehyde crosslinking step. In growing HeLa cells, both 3C (Chromosome Confirmation Capture) and MAR-Loop results show that the upstream region of the promoters and downstream transcription termination region are juxtaposed to each other. Therefore, we concluded that the anchorage of the IGS to

the nucleolar matrix could promote rDNA loop structures. Next, we performed a Chip-qPCR assay in serum stimulated HeLa cells and showed physical binding sites of Myc across the rDNA repeats (40kb). Moreover, we used Chip-loop (3C-Chip) assays and further determined the physical association of Myc in the loop organization of the IGS regions.

In general, genes can epigenetically silenced by DNA methylation as well as Histone modifications. In mouse cells, rRNA transcription can be regulated through a single CpG methylation at -133 in their Upstream Core Elements (UCE) [162, 163]. This single methylation in the UCE site can inhibit UBF binding to the rDNA and epigenetically abrogate Pol I pre-initiation complex formation [162, 163]. We used TGR1 cells (rat cell line) to further determine rDNA methylation status of matrix-attached regions, by methylation sensitive enzyme digestion, respectively HpaII and MspI. Our results illustrate that 80% of matrix associated rDNA is cleaved by HpaII, in comparison with 20% cleavage in non-matrix associated rDNAs, suggesting that matrix associated rRNA genes are hypo-methylated and more available for transcription activation.

7.3 Paper III: lymphoma-associated gene expression changes in an inducible model of Myc-driven B-cell lymphoma

In our previous study (Paper I), we illustrated that BL mutations could differentially change the c-Myc protein disorder profile. Moreover, previous functional and microarray studies in Rat fibroblasts revealed that adaptive cancer mutations could differentially affect Myc functions, compared to both the wild type Myc and other mutants. To better understand Myc's role in Burkitt's lymphoma, and how this is augmented by BL adaptive mutation we transduced primary B-cells with lentivirus constructs that allow regulatable conversion of the cells to lymphoma-like cells that depend on expression of WTMyc, T58A or T58I.

Over-expression of the c-Myc, in primary B-cells can induce apoptosis, and additional genetic elements are required for their cellular transformation [40]. To obtain Myc-driven lymphoma-like cell lines, we over-expressed c-Myc and inhibited the intrinsic and p53 apoptosis pathways, respectively by retroviral transduction of c-Myc, BCLXL and BMI1 expressing constructs. The transduced cells were pre-dominantly B-cells because they expressed CD19 and CD45R receptors but not the CD3 T-cell marker.

Unlike previous studies, in our lymphoma-like cell system we progressively increased Myc expression levels by addition of increasing doxycycline levels in the culture medium and growing cells for 48 hours, which was followed by different experiments including flow cytometry and RNA sequencing. To simplify our analysis steps, we first compared 7 different doxycycline dose dependent expression levels of Myc in each cell system (WTMyc, T58AMyc or T58IMyc), and next we compared the data between WTMyc vs mutants or T58A vs T58I, respectively.

We first looked at Myc's role in cell cycle progression, by analyzing cell cycle characteristics in response to the progressive increased Myc levels in all cell systems. Our result indicates that cell cycle phenotype of WTMyc and BL-AM (T58AMyc and T58IMyc) has a progressive increase in the proportion of cells in S-phase as Myc expression levels increase (0 to 1000 µg/ml). Moreover, cell populations representing the G2/M phase, exhibit lower changes in response to increased Myc levels. However, T58AMyc and T58IMyc expressing cells have a lower proportion of cells in G2M compared to cells expressing WTMyc.

To understand Myc and the phenotypic changes induced by its progressively increased expression during development of the lymphoma phenotype, as well as how these changes are augmented by adaptive BL mutations, we used inducible lymphoma-like cell systems to study their global gene expression by RNA sequencing. We first normalized count data, by adjusting to library size and gene length to produce TPM (Transcript Per Million) values. In order to choose the statistical methods and analyze our 3 dimensional data matrix (transcript counts vs Myc levels vs cell-lines), we first considered that the RNA sequencing count data had a skewed distribution. Second, we have multifactorial conditions in our data matrix. As previously described by other groups, the analytical approaches for multifactorial RNA-Seq data sets are sensitive to the normalization and statistical approaches, where as they report generalized linear models using non-binominal distribution are more accurate to detect differentially expressed genes (DEG) in multifactorial RNA-seq read count datasets [164, 165]. Using generalized linear models we identified 4443 DEGs commonly changed in response to the progressive increase of the Myc levels in WTMyc cells as well as BL-AM. Moreover, we identified 543 DEG, which were deregulated only in BL-AM, suggesting that BL-Associated Mutations can deregulate a different subset of the Myc target genes in lymphoma-like systems. Interestingly, we found two DEG subsets that were commonly expressed in wild type and either of the mutants; WT&T58A (n=553) and WT&T58I (n=1062). However, we need to perform further analyses to find their biological roles.

So far, we chose 6067 DEG in WTMyc expressing cells to visualize Myc transcriptional activities in WTMyc as well as BL-AM cell systems. Using the MaSigPro package in R we obtained 15 DEG model profiles, which mostly represent the different regulatory patterns regarding up and down regulated genes for all three Myc cell systems. Cells expressing T58I Myc proteins show the highest transcriptional levels in clusters 4, 8, 10 and 12 compared to WTMyc and T58AMyc cell systems. In contrast, cells that expressed T58A Myc proteins show the lowest values for down-regulated genes in cluster 13, where as in cluster 15, cells expressing T58A have higher expression values compared to the other cells.

Subsequently, we looked at the biological function of DEG in each defined cluster, by testing for enrichment of Gene Ontology (GO terms) using ClusterProfiler package (v3.6) in R. Interestingly, biological activity related to the ribosome biogenesis was enriched in cluster 9 and 11, and in cluster 9 both of the BL-AM and in cluster 11, the T58AMyc mutant exhibited

higher transcriptional levels compare to the WTMyc cell line. Moreover, in cluster profiles 10, 12 and 13 T58IMyc expressing cells, exhibited lower down regulation for the DEG associated with negative regulation of NF-KppaB activity, autophagy, cell adhesion, cell assembly, T-cell proliferation, histone and chromatin modifications, respectively, compared to the WTMyc and T58AMyc expressing cells.

8 DISCUSSION AND FUTURE PERSPECTIVES

Paper I: Phylogenetic and molecular evolution studies of Myc family proteins

Transcription Factors (TF) are highly disordered proteins, proteome wide analysis revealed that from 94 to 82 % of TFs have extended intrinsic disorder regions (IDR), which include many transcriptional activation domains [77]. This characteristic appears to be important for different functional roles, including interaction with and recruitment of different structurally unrelated partner proteins [75].

Intrinsic disorder prediction algorithms can reliably predict the propensity of the protein to be disordered for each amino acid residue along the length of a protein. By analogy and partial similarity to structured protein situations, it could be concluded that disordered pattern might be conserved in Myc family proteins, although the amino acid sequences are highly divergent. Thus, evolutionary conservation of the IDR patterns could be a useful tool for the phylogenetic classification of Myc proteins in mammals. In particular, we compared different dendograms using disorder data predicted by different algorithms, which showed positive correlations for all dendograms as well as some clear differences between different IDR predictor algorithms. Importantly, protein disorder prediction should be regarded as a complementary approach beside sequence based methods, for measuring Myc family protein relationships, and this IDR based method should not be considered as a replacement for DNA and protein approaches.

Moreover, further studies will be necessary in order to understand different parameters that influence the use of intrinsic disorder data and their relationship to identifying characteristics of proteins. In particular it will be of interest to analyze the evolution of different transcription factors as well as other proteins using both sequence based and the intrinsic disorder profiling methods. Indeed, additional approaches implemented using IUPred [166, 167](ANCHOR) and PONDR®VLXT [156, 168] (alpha-MoRF) can identify short ordered regions within the disorder region, as possible binding domains in the disordered regions. Comparing this data beside sequence based conserved domains of TFs could be informative for understanding rapid evolution of TFs and the functional characteristics of their conserved binding domains.

Alternatively, adaptive mutations might change biological functionality of a protein by changing the disorder profile of the disordered proteins, and therefore may influence the protein propensity for interacting with other partner proteins. In the BL-associated mutation context, it is interesting to report that functionally relevant c-Myc mutations can differentially change the local intrinsic disorder of the c-Myc protein.

Paper II: Functional analysis of Myc family proteins in mammalian cells

Yeast studies revealed that genes transcribed by RNA polymerase II form loop structures at the juxtaposition of the upstream and downstream regions of genes [169]. Thus, this spatial architecture of genes is thought to put the transcribing RNA polymerase in close proximity to promoter regions in order to facilitate re-initiation of transcription upon termination of a previous transcript [170], thus making a more efficient transcription hub for rDNAs. Previously, our group suggested a model in which c-Myc can bind to the rDNA and induce such gene looping in growing cells[121].

In paper II, we characterized and report a new role of c-Myc in rDNA transcription of mammalian cells. Firstly, we studied the nucleolus compartment and showed the role of c-Myc in rDNA transcription and its role of mediating nuclear matrix/scaffold attachment of rDNA. This role of the c-Myc protein in the nucleolus has not been reported before nor have similar roles of c-Myc been described in non-nucleolar contexts. Using Chip-3C assay we revealed that c-Myc can physically mediate rDNA attachment to the matrix, coupling with the MAR-methylation assay, which implies these matrix-attached rRNA genes are Hypo-methylated in the upstream core promoter and thereby more accessible for transcription.

Further investigation is needed to determine how Myc interaction with its co-factors and other nucleolar proteins, including UBF and SL1 can mediate this spatial organization of the mammalian rDNA.

Paper III: Lymphoma-associated gene expression changes in an inducible model of Myc-driven B-cell Lymphoma

We have been interested to understand Myc's roles in lymphoma development and how Burkitt's lymphoma associated mutations affect the Myc protein and augment Myc's roles. The most prominent recent in vitro studies to date have used inducible Myc expression in the P493-6 (B-cell) and U2OS (osteosarcoma) cell lines [171, 172]. In addition, fibroblast cell lines have frequently been used for detailed analysis of the role played by wild type Myc [172] and Burkitt Lymphoma Myc mutations [32]. Increase in ribosome biogenesis is a key factor for cancer cell growth and proliferation, during which new protein synthesis is critical for cancer cell proliferation. In this study, we found a significant enrichment of gene ontology terms for biological processes related to ribosome biogenesis, in differentially expressed gene clusters showing increased gene expression as the function of progressively increased expression of Myc. Our data is partially inconsistent with the studies of Myc in osteosarcomas suggesting that Myc can play different biological roles in different types of cancer cells. However, at the level of cell phenotype, our findings are consistent with the other studies of Myc inducible systems, where Myc over-expression was associated with increased cell size and entry of cells into the S-phase of the cell cycle [37, 172]. To further understand and address how Burkitt's lymphoma associated mutations contribute to lymphoma

development we made lymphoma-like transduced cells for the MycT58A and MycT58I mutations as well as MycWT. In comparison, a previous study on Myc and BL mutations in rat fibroblast-derived cell lines [32], we have found that in both data sets WTMyc and Burkitt's lymphoma associated mutations regulate a major subset of common genes, as well as different subsets with specific DEG for Burkitt's lymphoma associated mutations. In our results, deregulation of this major common gene subset (4443 genes) suggests that the mutations affecting Myc protein do not affect all Myc target genes. In addition, we observed that each BL-AM regulates a subset of genes similar to WTMyc but distinct from the other BL-AM. This contrasts with the rat fibroblast study that showed that BL-AM (T58I and E39D) had a very similar set of target genes, which is distinct from the target genes of wild type Myc. We need to in the future analyze the T58A (553) and T58I (1062) distinct subset of genes in order to find differences in their regulatory pathways.

Moreover, we need to compare our GO data set with the previous works done in Myc driven cells. Since, mutant Myc has a different half-life compared to wild type and for different mutants we need to further relate our data set to the actual protein levels of WTMyc and BL-AM in the studied cells.

To further understand the influence of BL mutations, which change the intrinsic disorder profile of the Myc protein we need to study their effect on the propensity of Myc to interact with and recruit partner proteins. We need to explore on the proteome level and analyze Myc interacting partners by different methods including Myc Immunoprecipitation and mass spectrometry.

However, major aspects of how Myc and BL-AM affect ribosome biogenesis remain unclear. By using transduced cells from this study and our previous knowledge of Myc's role in MAR and rDNA loop induction, it will be interesting to determine, how Myc and other nucleolar proteins can affect rDNA transcription during lymphoma development during the active transcription stage via different structural and epigenetic mechanisms.

9 ACKNOWLEDGEMENTS

I would like to take the opportunity to thank and express my deepest gratitude to all of you who made this dissertation possible.

Tony, My main supervisor; I would like to start my acknowledgment with two words Great HUMAN this is the simplest way to describe you, and express my deep appreciation and gratitude to have you as a mentor. Thanks for all the support, help, trust and guidance's for me to become a better scientist as well as a person. I enjoyed our discussions and the knowledge you had shared with me over the years. Your curiosity and broad knowledge in different fields have always inspired me, and also your life long dedication in research, teaching as well as your family.

Kathrin, words can not expressed how grateful I am to have you as a friend and mentor. Through these years you always listened to me, never judged, and helped me to understand and overcome the problems and go ahead.

Hamed, ☺☺ the best part of working in NOVUM was having my best friend around in the lab. You have always been supporting me, pushing me to do the right things and helping me in both science and life.

Alf, it was a pleasure collaborating with you. Your advice and comment were always appreciated.

Mina, thank you so much for all the help, support, valuable advice and feed backs.

Mohsen, you are one the most brilliant scientist that I have ever known. A knowledgeable and dedicated person. I always appreciate your technical advices for developing methods. But in reality, I LIKE you Buddy as one of my best friends. Thank you agha Mohsen vase hame mohabataton as well as your pure friendship.

Samir, you are always inspiring me as a successful non-boring scientist, with a great heart. I always appreciate to have discussions with a talented scientist like you.

Iyadh, I always enjoyed our short talks in the corridor. You are a brilliant scientist and I really appreciate our discussions, where you shared your technical and scientific knowledge with me.

Behnam, I appreciate your friendship and the time you always put for me. Your help and advices were always useful in my life. As well as, my name is amir and not hamid agha behnam.

Helmi, thank you for every day talk, being generous to the entire faculty and students. I missed you in our group. **Laia**, you were the best addition to our group, a positive person that I appreciate sharing the office with a person that always smiles. **Gustave**, my bioinformatics

mentor, I appreciate our collaboration and what I have learned from you, it was nice sharing the office with you. From cycling and photography till bioinformatics and science, I am amazed by how knowledgeable and thoughtful you are. **YongTao** = fun, beautiful, super social person with a positive attitude in life, I always enjoy our coffee times and parties.

Chiou-Nan, you are a hard worker, dedicated scientist I wish you all the best.

I would like to express my sincere appreciation and gratitude to all my colleagues in Ted and Samir groups; I enjoyed being a member of the groups and sharing the lab or office with you.

Edvard Smite, it was an honor to know you as a scientist and joining your group. **Karin, Beklem, Anna, Dara, Giulia, Fadwa, Olof, Dhanu, Oskar, Tiia, Helena, Joel, Manuela, Osama, Oscar, Yesid, Oksana, Velademir, Rula, Han, Chin, Min, Taavi, Ulrika, Barance, Jeremy, Maria, Andre, Caroline, Narmin, Karim, Tea.**

Emelie, Caroline and Hanna, thank you so much, you were always kind, helpful and friendly. **BEA** core facility, especially **Patrick** and **Fredrik**, who I enjoyed talking about the sequencing as well as the motorbikes.

I want to thank my Persian friends; I always enjoy spending time with you all.

Maryam, Thank you for helping me and supporting me from the time of my Master studies. I can't believe it, I knew you for almost 8 years, which I appreciate your friendship, I cannot believe that I know you for almost 8 years, and appreciate your friendship. **Ali**, you were my first classmate in Sweden, which we both continue our scientific career; I admire your great personality and our discussions at fika times. **Marjan**, I always admire your strong personality and I like our talks about sports. You are a brave girl with a big heart. **Feri**, our friendship started with the color festival in spring and remained colorful through all these years, thank you buddy for supporting me. **Reza**, you are a well-organized person and a trustworthy friend that always helped me when I needed it. **Hassan**, you are always motivating me, I enjoy our conversation from bioinformatics to Netflix, human evolution and so on. Somehow we always have fun combined with the science and "jok va digar hich". **Parisa**, you are a good friend with a calm attitude. **Yasaman** a motivated person which I had enjoyed talking with you. **Sara**, Smart Persian Professor, I always enjoyed having coffee with you and teaching you some new Persian words. And special thanks and have the best wishes for my good friends **Shabnam khanum raeis bozorg, Aida, Maryam, Yaser, Mohammad and Maryam khanum.**

Rock climbing with the scientist was another awesome experience, **Ola**; you are a sweet and kind friend, which is always ready to step up for a new challenge. **Angelina**, I am so lucky to have you as a friend, and I always enjoy climbing with you. **Robi**, my dear trustworthy friend, we always had great discussions when we belayed others.

Haleh, you had taught me the most beautiful things in my life, I am always grateful that I know you and had spent my time with you. You always believe in me and give me the motivation to go forward, you have been always beside me no matter where I am.

My parents, you always believed and supported me, I do appreciate having knowledgeable parents. You showed me how to be a human, love and help the others. You are my heroes.

Maryam, My talented sister, you are the best example of an independent successful woman, I am so proud of you. We do not speak often but you are always around the corner helping me. **Alireza**, how lucky I am to have you beside me 😊. it was one of the most beautiful parts of my life seeing you growing up, and developing as a person, you are a gentleman with a great Soul, a hard working scientist with a great passion for nature. **Nouradin**, My dearest uncle, I start science in your lab and I miss you the most in my dissertation” doseton daram amo jan”. **Clayton**, My big brother, I always enjoyed spending time with you. **Maryam**, you are a great person with a nice smile and a great future ahead of you.

10 REFERENCES

1. Sheiness, D. and J.M. Bishop, *DNA and RNA from uninfected vertebrate cells contain nucleotide sequences related to the putative transforming gene of avian myelocytomatosis virus*. J Virol, 1979. **31**(2): p. 514-21.
2. Dalla-Favera, R., et al., *Human c-myc onc gene is located on the region of chromosome 8 that is translocated in Burkitt lymphoma cells*. Proc Natl Acad Sci U S A, 1982. **79**(24): p. 7824-7.
3. Nesbit, C.E., J.M. Tersak, and E.V. Prochownik, *MYC oncogenes and human neoplastic disease*. Oncogene, 1999. **18**(19): p. 3004-16.
4. Luscher, B. and J. Vervoorts, *Regulation of gene transcription by the oncoprotein MYC*. Gene, 2012. **494**(2): p. 145-60.
5. Eilers, M. and R.N. Eisenman, *Myc's broad reach*. Genes Dev, 2008. **22**(20): p. 2755-66.
6. Adhikary, S. and M. Eilers, *Transcriptional regulation and transformation by Myc proteins*. Nat Rev Mol Cell Biol, 2005. **6**(8): p. 635-45.
7. Grandori, C., et al., *The Myc/Max/Mad network and the transcriptional control of cell behavior*. Annu Rev Cell Dev Biol, 2000. **16**: p. 653-99.
8. Albiñ, A., J.I. Johnsen, and M.A. Henriksson, *MYC in oncogenesis and as a target for cancer therapies*. Adv Cancer Res, 2010. **107**: p. 163-224.
9. Dang, C.V., *MYC on the path to cancer*. Cell, 2012. **149**(1): p. 22-35.
10. Tansey, W.P., *Mammalian MYC Proteins and Cancer*. New Journal of Science, 2014. **2014**: p. 27.
11. Larsson, L.G. and M.A. Henriksson, *The Yin and Yang functions of the Myc oncoprotein in cancer development and as targets for therapy*. Exp Cell Res, 2010. **316**(8): p. 1429-37.
12. Luscher, B. and L.G. Larsson, *The basic region/helix-loop-helix/leucine zipper domain of Myc proto-oncoproteins: function and regulation*. Oncogene, 1999. **18**(19): p. 2955-66.
13. Henriksson, M. and B. Lüscher, *Proteins of the Myc Network: Essential Regulators of Cell Growth and Differentiation*, in *Advances in Cancer Research*, G.F. Vande Woude and G. Klein, Editors. 1996, Academic Press. p. 109-182.
14. Stanton, B.R., et al., *Loss of N-myc function results in embryonic lethality and failure of the epithelial component of the embryo to develop*. Genes Dev, 1992. **6**(12a): p. 2235-47.
15. Davis, A.C., et al., *A null c-myc mutation causes lethality before 10.5 days of gestation in homozygotes and reduced fertility in heterozygous female mice*. Genes Dev, 1993. **7**(4): p. 671-82.

16. West, M.J., M. Stoneley, and A.E. Willis, *Translational induction of the c-myc oncogene via activation of the FRAP/TOR signalling pathway*. *Oncogene*, 1998. **17**(6): p. 769-80.
17. Roussel, M.F., et al., *Myc rescue of a mutant CSF-1 receptor impaired in mitogenic signalling*. *Nature*, 1991. **353**: p. 361.
18. Iritani, B.M. and R.N. Eisenman, *c-Myc enhances protein synthesis and cell size during B lymphocyte development*. *Proc Natl Acad Sci U S A*, 1999. **96**(23): p. 13180-5.
19. Gomez-Roman, N., et al., *Direct activation of RNA polymerase III transcription by c-Myc*. *Nature*, 2003. **421**(6920): p. 290-4.
20. Arabi, A., et al., *c-Myc associates with ribosomal DNA and activates RNA polymerase I transcription*. *Nat Cell Biol*, 2005. **7**(3): p. 303-10.
21. Grandori, C., et al., *c-Myc binds to human ribosomal DNA and stimulates transcription of rRNA genes by RNA polymerase I*. *Nat Cell Biol*, 2005. **7**(3): p. 311-8.
22. McMahon, S.B., M.A. Wood, and M.D. Cole, *The essential cofactor TRRAP recruits the histone acetyltransferase hGCN5 to c-Myc*. *Mol Cell Biol*, 2000. **20**(2): p. 556-62.
23. Nikiforov, M.A., et al., *TRRAP-dependent and TRRAP-independent transcriptional activation by Myc family oncoproteins*. *Mol Cell Biol*, 2002. **22**(14): p. 5054-63.
24. Liu, X., et al., *c-Myc transformation domain recruits the human STAGA complex and requires TRRAP and GCN5 acetylase activity for transcription activation*. *J Biol Chem*, 2003. **278**(22): p. 20405-12.
25. Doyon, Y., et al., *Structural and functional conservation of the NuA4 histone acetyltransferase complex from yeast to humans*. *Mol Cell Biol*, 2004. **24**(5): p. 1884-96.
26. Fladvad, M., et al., *N and C-terminal sub-regions in the c-Myc transactivation region and their joint role in creating versatility in folding and binding*. *J Mol Biol*, 2005. **346**(1): p. 175-89.
27. Martinato, F., et al., *Analysis of Myc-induced histone modifications on target chromatin*. *PLoS One*, 2008. **3**(11): p. e3650.
28. Oster, S.K., et al., *Functional analysis of the N-terminal domain of the Myc oncoprotein*. *Oncogene*, 2003. **22**(13): p. 1998-2010.
29. Hydbring, P., A. Castell, and L.G. Larsson, *MYC Modulation around the CDK2/p27/SKP2 Axis*. *Genes (Basel)*, 2017. **8**(7).
30. Welcker, M., et al., *The Fbw7 tumor suppressor regulates glycogen synthase kinase 3 phosphorylation-dependent c-Myc protein degradation*. *Proc Natl Acad Sci U S A*, 2004. **101**(24): p. 9085-90.
31. Farrell, A.S. and R.C. Sears, *MYC degradation*. *Cold Spring Harb Perspect Med*, 2014. **4**(3).

32. Cowling, V.H., S.A. Turner, and M.D. Cole, *Burkitt's lymphoma-associated c-Myc mutations converge on a dramatically altered target gene response and implicate Ncl5a/Nop56 in oncogenesis*. *Oncogene*, 2014. **33**(27): p. 3519-27.
33. Chakraborty, A.A., et al., *A common functional consequence of tumor-derived mutations within c-MYC*. 2015. **34**(18): p. 2406-9.
34. Xu-Monette, Z.Y., et al., *Clinical and Biologic Significance of MYC Genetic Mutations in De Novo Diffuse Large B-cell Lymphoma*. *Clin Cancer Res*, 2016. **22**(14): p. 3593-605.
35. Chang, D.W., et al., *The c-Myc transactivation domain is a direct modulator of apoptotic versus proliferative signals*. *Mol Cell Biol*, 2000. **20**(12): p. 4309-19.
36. Cowling, V.H., et al., *A conserved Myc protein domain, MBIV, regulates DNA binding, apoptosis, transformation, and G2 arrest*. *Mol Cell Biol*, 2006. **26**(11): p. 4226-39.
37. Walz, S., et al., *Activation and repression by oncogenic MYC shape tumour-specific gene expression profiles*. *Nature*, 2014. **511**(7510): p. 483-7.
38. Mateyak, M.K., et al., *Phenotypes of c-Myc-deficient rat fibroblasts isolated by targeted homologous recombination*. *Cell Growth Differ*, 1997. **8**(10): p. 1039-48.
39. Pelengaris, S., M. Khan, and G. Evan, *c-MYC: more than just a matter of life and death*. *Nat Rev Cancer*, 2002. **2**(10): p. 764-76.
40. Hogstrand, k. and A. Grandien, *MYC-driven tumor transformation of mature B cells requires inhibition of both intrinsic apoptosis and p53 activity*. Manuscript send for *European Journal of Immunology*, 2018.
41. Wagner, A.J., J.M. Kokontis, and N. Hay, *Myc-mediated apoptosis requires wild-type p53 in a manner independent of cell cycle arrest and the ability of p53 to induce p21waf1/cip1*. *Genes Dev*, 1994. **8**(23): p. 2817-30.
42. Zindy, F., et al., *Myc signaling via the ARF tumor suppressor regulates p53-dependent apoptosis and immortalization*. *Genes Dev*, 1998. **12**(15): p. 2424-33.
43. Seoane, J., H.V. Le, and J. Massague, *Myc suppression of the p21(Cip1) Cdk inhibitor influences the outcome of the p53 response to DNA damage*. *Nature*, 2002. **419**(6908): p. 729-34.
44. Wu, S., et al., *Myc represses differentiation-induced p21CIP1 expression via Miz-1-dependent interaction with the p21 core promoter*. *Oncogene*, 2003. **22**(3): p. 351-60.
45. Harris, A.W., et al., *The E mu-myc transgenic mouse. A model for high-incidence spontaneous lymphoma and leukemia of early B cells*. *J Exp Med*, 1988. **167**(2): p. 353-71.
46. Muller, J.R., et al., *Persistence of immunoglobulin heavy chain/c-myc recombination-positive lymphocyte clones in the blood of human immunodeficiency virus-infected homosexual men*. *Proc Natl Acad Sci U S A*, 1995. **92**(14): p. 6577-81.

47. Eischen, C.M., et al., *Disruption of the ARF-Mdm2-p53 tumor suppressor pathway in Myc-induced lymphomagenesis*. Genes Dev, 1999. **13**(20): p. 2658-69.
48. Jacobs, J.J., et al., *Bmi-1 collaborates with c-Myc in tumorigenesis by inhibiting c-Myc-induced apoptosis via INK4a/ARF*. Genes Dev, 1999. **13**(20): p. 2678-90.
49. Schmitt, C.A., et al., *INK4a/ARF mutations accelerate lymphomagenesis and promote chemoresistance by disabling p53*. Genes Dev, 1999. **13**(20): p. 2670-7.
50. Swanson, P.J., et al., *Fatal acute lymphoblastic leukemia in mice transgenic for B cell-restricted bcl-xL and c-myc*. J Immunol, 2004. **172**(11): p. 6684-91.
51. Campbell, K.J., et al., *Elevated Mcl-1 perturbs lymphopoiesis, promotes transformation of hematopoietic stem/progenitor cells, and enhances drug resistance*. Blood, 2010. **116**(17): p. 3197-207.
52. Land, H., L.F. Parada, and R.A. Weinberg, *Tumorigenic conversion of primary embryo fibroblasts requires at least two cooperating oncogenes*. Nature, 1983. **304**(5927): p. 596-602.
53. Bissonnette, R.P., et al., *Apoptotic cell death induced by c-myc is inhibited by bcl-2*. Nature, 1992. **359**(6395): p. 552-4.
54. Arvanitis, C. and D.W. Felsher, *Conditional transgenic models define how MYC initiates and maintains tumorigenesis*. Semin Cancer Biol, 2006. **16**(4): p. 313-7.
55. Deshaies, R.J., *SCF and Cullin/Ring H2-based ubiquitin ligases*. Annu Rev Cell Dev Biol, 1999. **15**: p. 435-67.
56. Welcker, M., et al., *A nucleolar isoform of the Fbw7 ubiquitin ligase regulates c-Myc and cell size*. Curr Biol, 2004. **14**(20): p. 1852-7.
57. Yada, M., et al., *Phosphorylation-dependent degradation of c-Myc is mediated by the F-box protein Fbw7*. Embo j, 2004. **23**(10): p. 2116-25.
58. Farrell, A.S., et al., *Pin1 regulates the dynamics of c-Myc DNA binding to facilitate target gene regulation and oncogenesis*. Mol Cell Biol, 2013. **33**(15): p. 2930-49.
59. Sanchez-Arevalo Lobo, V.J., et al., *Dual regulation of Myc by Abl*. Oncogene, 2013. **32**(45): p. 5261-71.
60. Gregory, M.A., Y. Qi, and S.R. Hann, *Phosphorylation by glycogen synthase kinase-3 controls c-myc proteolysis and subnuclear localization*. J Biol Chem, 2003. **278**(51): p. 51606-12.
61. Yeh, E., et al., *A signalling pathway controlling c-Myc degradation that impacts oncogenic transformation of human cells*. Nat Cell Biol, 2004. **6**(4): p. 308-18.
62. Kim, S.Y., et al., *Skp2 regulates Myc protein stability and activity*. Mol Cell, 2003. **11**(5): p. 1177-88.
63. von der Lehr, N., et al., *The F-box protein Skp2 participates in c-Myc proteasomal degradation and acts as a cofactor for c-Myc-regulated transcription*. Mol Cell, 2003. **11**(5): p. 1189-200.
64. Vervoorts, J., et al., *Stimulation of c-MYC transcriptional activity and acetylation by recruitment of the cofactor CBP*. EMBO Rep, 2003. **4**(5): p. 484-90.

65. Patel, J.H., et al., *The c-MYC oncoprotein is a substrate of the acetyltransferases hGCN5/PCAF and TIP60*. Mol Cell Biol, 2004. **24**(24): p. 10826-34.
66. Zhang, K., F. Faiola, and E. Martinez, *Six lysine residues on c-Myc are direct substrates for acetylation by p300*. Biochem Biophys Res Commun, 2005. **336**(1): p. 274-80.
67. Yuan, J., K. Minter-Dykhouse, and Z. Lou, *A c-Myc-SIRT1 feedback loop regulates cell growth and transformation*. J Cell Biol, 2009. **185**(2): p. 203-11.
68. Gallant, P., *Myc/Max/Mad in invertebrates: the evolution of the Max network*. Curr Top Microbiol Immunol, 2006. **302**: p. 235-53.
69. Marandel, L., et al., *Evolutionary history of c-myc in teleosts and characterization of the duplicated c-myca genes in goldfish embryos*. Mol Reprod Dev, 2012. **79**(2): p. 85-96.
70. King, N., et al., *The genome of the choanoflagellate Monosiga brevicollis and the origin of metazoans*. Nature, 2008. **451**: p. 783.
71. Young, S.L., et al., *Premetazoan ancestry of the Myc-Max network*. Mol Biol Evol, 2011. **28**(10): p. 2961-71.
72. McFerrin, L.G. and W.R. Atchley, *Evolution of the Max and Mlx networks in animals*. Genome Biol Evol, 2011. **3**: p. 915-37.
73. Skinner, M.K., et al., *Basic helix-loop-helix transcription factor gene family phylogenetics and nomenclature*. Differentiation, 2010. **80**(1): p. 1-8.
74. Johnston, L.A., et al., *Drosophila myc regulates cellular growth during development*. Cell, 1999. **98**(6): p. 779-90.
75. Staby, L., et al., *Eukaryotic transcription factors: paradigms of protein intrinsic disorder*. Biochem J, 2017. **474**(15): p. 2509-2532.
76. Banerjee, S., S. Chakraborty, and R.K. De, *Deciphering the cause of evolutionary variance within intrinsically disordered regions in human proteins*. J Biomol Struct Dyn, 2017. **35**(2): p. 233-249.
77. Liu, J., et al., *Intrinsic disorder in transcription factors*. Biochemistry, 2006. **45**(22): p. 6873-88.
78. Garner, E., et al., *Predicting Binding Regions within Disordered Proteins*. Genome Inform Ser Workshop Genome Inform, 1999. **10**: p. 41-50.
79. Oldfield, C.J., et al., *Coupled folding and binding with alpha-helix-forming molecular recognition elements*. Biochemistry, 2005. **44**(37): p. 12454-70.
80. Ota, H. and S. Fukuchi, *Sequence conservation of protein binding segments in intrinsically disordered regions*. Biochem Biophys Res Commun, 2017. **494**(3-4): p. 602-607.
81. Campbell, K.J. and R.J. White, *MYC regulation of cell growth through control of transcription by RNA polymerases I and III*. Cold Spring Harb Perspect Med, 2014. **4**(5).

82. Rudra, D., Y. Zhao, and J.R. Warner, *Central role of Ifh1p-Fhl1p interaction in the synthesis of yeast ribosomal proteins*. *Embo j*, 2005. **24**(3): p. 533-42.
83. Andersen, J.S., et al., *Nucleolar proteome dynamics*. *Nature*, 2005. **433**(7021): p. 77-83.
84. Boisvert, F.M., et al., *The multifunctional nucleolus*. *Nat Rev Mol Cell Biol*, 2007. **8**(7): p. 574-85.
85. Gonzalez, I.L. and J.E. Sylvester, *Complete sequence of the 43-kb human ribosomal DNA repeat: analysis of the intergenic spacer*. *Genomics*, 1995. **27**(2): p. 320-8.
86. Ivan, R., et al., *The nucleolus and transcription of ribosomal genes*. *Biology of the Cell*, 2004. **96**(8): p. 579-594.
87. Miller, O.L., Jr. and B.R. Beatty, *Visualization of nucleolar genes*. *Science*, 1969. **164**(3882): p. 955-7.
88. Henderson, A.S., D. Warburton, and K.C. Atwood, *Location of ribosomal DNA in the human chromosome complement*. *Proc Natl Acad Sci U S A*, 1972. **69**(11): p. 3394-8.
89. Gonzalez, I.L. and J.E. Sylvester, *Beyond ribosomal DNA: on towards the telomere*. *Chromosoma*, 1997. **105**(7-8): p. 431-7.
90. Wright, J.E., et al., *A role for upstream binding factor in organizing ribosomal gene chromatin*. *Biochem Soc Symp*, 2006(73): p. 77-84.
91. Maggi, L.B., Jr. and J.D. Weber, *Nucleolar adaptation in human cancer*. *Cancer Invest*, 2005. **23**(7): p. 599-608.
92. Hernandez-Verdun, D., *The nucleolus: a model for the organization of nuclear functions*. *Histochem Cell Biol*, 2006. **126**(2): p. 135-48.
93. Hozak, P., et al., *Site of transcription of ribosomal RNA and intranucleolar structure in HeLa cells*. *J Cell Sci*, 1994. **107 (Pt 2)**: p. 639-48.
94. Sirri, V., et al., *Nucleolus: the fascinating nuclear body*. *Histochem Cell Biol*, 2008. **129**(1): p. 13-31.
95. Russell, J. and J.C. Zomerdijk, *RNA-polymerase-I-directed rDNA transcription, life and works*. *Trends Biochem Sci*, 2005. **30**(2): p. 87-96.
96. Liao, M.C. and R.P. Perry, *Ribosome precursor particles in nucleoli*. *J Cell Biol*, 1969. **42**(1): p. 272-83.
97. Berezney, R., et al., *The nuclear matrix: a structural milieu for genomic function*. *Int Rev Cytol*, 1995. **162a**: p. 1-65.
98. Brown, K.E., et al., *Association of transcriptionally silent genes with Ikaros complexes at centromeric heterochromatin*. *Cell*, 1997. **91**(6): p. 845-54.
99. Mehta, I.S., et al., *Alterations to nuclear architecture and genome behavior in senescent cells*. *Ann N Y Acad Sci*, 2007. **1100**: p. 250-63.
100. de Las Heras, J.I. and E.C. Schirmer, *The nuclear envelope and cancer: a diagnostic perspective and historical overview*. *Adv Exp Med Biol*, 2014. **773**: p. 5-26.

101. Fischer, A.H., *The diagnostic pathology of the nuclear envelope in human cancers*. Adv Exp Med Biol, 2014. **773**: p. 49-75.
102. Osborne, C.S., et al., *Active genes dynamically colocalize to shared sites of ongoing transcription*. Nat Genet, 2004. **36**(10): p. 1065-71.
103. Albiez, H., et al., *Chromatin domains and the interchromatin compartment form structurally defined and functionally interacting nuclear networks*. Chromosome Res, 2006. **14**(7): p. 707-33.
104. Lanctot, C., et al., *Dynamic genome architecture in the nuclear space: regulation of gene expression in three dimensions*. Nat Rev Genet, 2007. **8**(2): p. 104-15.
105. Bode, J., et al., *Transcriptional augmentation: modulation of gene expression by scaffold/matrix-attached regions (S/MAR elements)*. Crit Rev Eukaryot Gene Expr, 2000. **10**(1): p. 73-90.
106. Kreth, G., et al., *Chromatin structure and chromosome aberrations: modeling of damage induced by isotropic and localized irradiation*. Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis, 1998. **404**(1): p. 77-88.
107. Brown, K.E., *Chromatin folding and gene expression: new tools to reveal the spatial organization of genes*. Chromosome Res, 2003. **11**(5): p. 423-33.
108. Odenheimer, J., G. Kreth, and D.W. Heermann, *Dynamic simulation of active/inactive chromatin domains*. J Biol Phys, 2005. **31**(3-4): p. 351-63.
109. Linnemann, A.K., A.E. Platts, and S.A. Krawetz, *Differential nuclear scaffold/matrix attachment marks expressed genes*. Hum Mol Genet, 2009. **18**(4): p. 645-54.
110. Halweg, C., W.F. Thompson, and S. Spiker, *The rb7 matrix attachment region increases the likelihood and magnitude of transgene expression in tobacco cells: a flow cytometric study*. Plant Cell, 2005. **17**(2): p. 418-29.
111. Cheutin, T., et al., *Three-dimensional organization of active rRNA genes within the nucleolus*. J Cell Sci, 2002. **115**(Pt 16): p. 3297-307.
112. Boukhgalter, B., et al., *Characterization of a fission yeast subunit of an RNA polymerase I essential transcription initiation factor, SpRrn7h/TAFI68, that bridges yeast and mammals: association with SpRrn1Ih and the core ribosomal RNA gene promoter*. Gene, 2002. **291**(1): p. 187-201.
113. Bazett-Jones, D.P., et al., *Short-range DNA looping by the Xenopus HMG-box transcription factor, xUBF*. Science, 1994. **264**(5162): p. 1134-7.
114. Moorefield, B., E.A. Greene, and R.H. Reeder, *RNA polymerase I transcription factor Rrn3 is functionally conserved between yeast and human*. Proc Natl Acad Sci U S A, 2000. **97**(9): p. 4724-9.
115. Grummt, I., *Life on a planet of its own: regulation of RNA polymerase I transcription in the nucleolus*. Genes Dev, 2003. **17**(14): p. 1691-702.
116. White, R.J., *RNA polymerases I and III, non-coding RNAs and cancer*. Trends in Genetics, 2008. **24**(12): p. 622-629.

117. Schramm, L. and N. Hernandez, *Recruitment of RNA polymerase III to its target promoters*. Genes Dev, 2002. **16**(20): p. 2593-620.
118. Hannan, K.M., et al., *Dysregulation of RNA polymerase I transcription during disease*. Biochim Biophys Acta, 2013. **1829**(3-4): p. 342-60.
119. Grewal, S.S., et al., *Myc-dependent regulation of ribosomal RNA synthesis during Drosophila development*. Nat Cell Biol, 2005. **7**(3): p. 295-302.
120. Ji, H., et al., *Cell-type independent MYC target genes reveal a primordial signature involved in biomass accumulation*. PLoS One, 2011. **6**(10): p. e26057.
121. Shiue, C.N., R.G. Berkson, and A.P. Wright, *c-Myc induces changes in higher order rDNA structure on stimulation of quiescent cells*. Oncogene, 2009. **28**(16): p. 1833-42.
122. Poortinga, G., et al., *c-MYC coordinately regulates ribosomal gene chromatin remodeling and Pol I availability during granulocyte differentiation*. Nucleic Acids Res, 2011. **39**(8): p. 3267-81.
123. Rosenbloom, K.R., et al., *ENCODE data in the UCSC Genome Browser: year 5 update*. Nucleic Acids Res, 2013. **41**(Database issue): p. D56-63.
124. Kenneth, N.S., et al., *TRRAP and GCN5 are used by c-Myc to activate RNA polymerase III transcription*. Proc Natl Acad Sci U S A, 2007. **104**(38): p. 14917-22.
125. Goodfellow, S.J., et al., *Regulation of RNA polymerase III transcription during hypertrophic growth*. Embo j, 2006. **25**(7): p. 1522-33.
126. Lin, C.Y., et al., *Transcriptional amplification in tumor cells with elevated c-Myc*. Cell, 2012. **151**(1): p. 56-67.
127. Shukla, S.K. and V. Kumar, *Hepatitis B virus X protein and c-Myc cooperate in the upregulation of ribosome biogenesis and in cellular transformation*. Febs j, 2012. **279**(20): p. 3859-71.
128. Ernens, I., et al., *Hypoxic stress suppresses RNA polymerase III recruitment and tRNA gene transcription in cardiomyocytes*. Nucleic Acids Res, 2006. **34**(1): p. 286-94.
129. Walker HK , H.W., Hurst JW, *The White Blood Cell and Differential Count*. Clinical Methods: The History, Physical, and Laboratory Examinations. 3rd edition. Boston: Butterworths. 1990. chapter 153.
130. Appay, V., et al., *Phenotype and function of human T lymphocyte subsets: consensus and issues*. Cytometry A, 2008. **73**(11): p. 975-83.
131. LeBien, T.W. and T.F. Tedder, *B lymphocytes: how they develop and function*. Blood, 2008. **112**(5): p. 1570-80.
132. Dalla-Favera, R., *Lymphoid malignancies: many tumor types, many altered genes, many therapeutic challenges*. Blood, 2012. **122**(10): p. 3396-7.
133. Swerdlow, S.H., et al., *The 2016 revision of the World Health Organization classification of lymphoid neoplasms*. 2016. **127**(20): p. 2375-90.

134. Klein, L., et al., *Positive and negative selection of the T cell repertoire: what thymocytes see (and don't see)*. Nat Rev Immunol, 2014. **14**(6): p. 377-91.
135. Krueger, A., *Thymus Colonization: Who, How, How Many?* Arch Immunol Ther Exp (Warsz), 2018. **66**(2): p. 81-88.
136. Shevach, E.M., *Regulatory T cells in autoimmunity**. Annu Rev Immunol, 2000. **18**: p. 423-49.
137. Luckheeram, R.V., et al., *CD4(+)T cells: differentiation and functions*. Clin Dev Immunol, 2012. **2012**: p. 925135.
138. Li, Y.S., K. Hayakawa, and R.R. Hardy, *The regulated expression of B lineage associated genes during B cell differentiation in bone marrow and fetal liver*. J Exp Med, 1993. **178**(3): p. 951-60.
139. Tobon, G.J., J.H. Izquierdo, and C.A. Canas, *B lymphocytes: development, tolerance, and their role in autoimmunity-focus on systemic lupus erythematosus*. Autoimmune Dis, 2013. **2013**: p. 827254.
140. Janeway, C.A., Jr., J. Ron, and M.E. Katz, *The B cell is the initiating antigen-presenting cell in peripheral lymph nodes*. J Immunol, 1987. **138**(4): p. 1051-5.
141. Lund, F.E., *Cytokine-producing B lymphocytes-key regulators of immunity*. Curr Opin Immunol, 2008. **20**(3): p. 332-8.
142. Mauri, C. and A. Bosma, *Immune regulatory function of B cells*. Annu Rev Immunol, 2012. **30**: p. 221-41.
143. Burkitt, D. and G.T. O'Connor, *Malignant lymphoma in African children. I. A clinical syndrome*. Cancer, 1961. **14**: p. 258-69.
144. Bellan, C., et al., *Burkitt's lymphoma: new insights into molecular pathogenesis*. J Clin Pathol, 2003. **56**(3): p. 188-92.
145. Blum, K.A., G. Lozanski, and J.C. Byrd, *Adult Burkitt leukemia and lymphoma*. Blood, 2004. **104**(10): p. 3009-20.
146. Yustein, J.T. and C.V. Dang, *Biology and treatment of Burkitt's lymphoma*. Curr Opin Hematol, 2007. **14**(4): p. 375-81.
147. Ott, G., A. Rosenwald, and E. Campo, *Understanding MYC-driven aggressive B-cell lymphomas: pathogenesis and classification*. Blood, 2013. **122**(24): p. 3884-91.
148. Taub, R., et al., *Translocation of the c-myc gene into the immunoglobulin heavy chain locus in human Burkitt lymphoma and murine plasmacytoma cells*. Proc Natl Acad Sci U S A, 1982. **79**(24): p. 7837-41.
149. Ott, G., A. Rosenwald, and E. Campo, *Understanding MYC-driven aggressive B-cell lymphomas: pathogenesis and classification*. Hematology Am Soc Hematol Educ Program, 2013. **2013**: p. 575-83.
150. Meyer, N. and L.Z. Penn, *Reflecting on 25 years with MYC*. Nat Rev Cancer, 2008. **8**(12): p. 976-90.

151. Fiskvik, I., et al., *Combining MYC, BCL2 and TP53 gene and protein expression alterations improves risk stratification in diffuse large B-cell lymphoma*. Leuk Lymphoma, 2015. **56**(6): p. 1742-9.
152. Gebauer, N., et al., *TP53 mutations are frequent events in double-hit B-cell lymphomas with MYC and BCL2 but not MYC and BCL6 translocations*. Leuk Lymphoma, 2015. **56**(1): p. 179-85.
153. Mahani, A., J. Henriksson, and A.P. Wright, *Origins of Myc proteins--using intrinsic protein disorder to trace distant relatives*. PLoS One, 2013. **8**(9): p. e75057.
154. Langmead, B. and S.L. Salzberg, *Fast gapped-read alignment with Bowtie 2*. Nat Methods, 2012. **9**(4): p. 357-9.
155. Anders, S., P.T. Pyl, and W. Huber, *HTSeq--a Python framework to work with high-throughput sequencing data*. Bioinformatics, 2015. **31**(2): p. 166-9.
156. Romero, P., et al., *Sequence complexity of disordered protein*. Proteins, 2001. **42**(1): p. 38-48.
157. Hadley, W., *Ggplot2*. 2016, New York, NY: Springer Science+Business Media, LLC. pages cm.
158. Cockerill, P.N. and W.T. Garrard, *Chromosomal loop anchorage of the kappa immunoglobulin gene occurs next to the enhancer in a region containing topoisomerase II sites*. Cell, 1986. **44**(2): p. 273-82.
159. Fiorini, A., S. Gouveia Fde, and M.A. Fernandez, *Scaffold/Matrix Attachment Regions and intrinsic DNA curvature*. Biochemistry (Mosc), 2006. **71**(5): p. 481-8.
160. Shiomi, Y., et al., *Proteins and RNA in mouse L cell core nucleoli and nucleolar matrix*. Biochemistry, 1986. **25**(19): p. 5745-51.
161. Grdovic, N., M. Vidakovic, and G. Poznanovic, *Binding of a 23 kD endonuclease to the rat liver nuclear matrix*. Gen Physiol Biophys, 2005. **24**(1): p. 99-111.
162. Santoro, R. and I. Grummt, *Molecular mechanisms mediating methylation-dependent silencing of ribosomal gene transcription*. Mol Cell, 2001. **8**(3): p. 719-25.
163. Santoro, R. and I. Grummt, *Epigenetic mechanism of rRNA gene silencing: temporal order of NoRC-mediated histone modification, chromatin remodeling, and DNA methylation*. Mol Cell Biol, 2005. **25**(7): p. 2539-46.
164. McCarthy, D.J., Y. Chen, and G.K. Smyth, *Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation*. Nucleic Acids Res, 2012. **40**(10): p. 4288-97.
165. Lin, Y., et al., *Comparison of normalization and differential expression analyses using RNA-Seq data from 726 individual Drosophila melanogaster*. BMC Genomics, 2016. **17**: p. 28.
166. Dosztanyi, Z., et al., *IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content*. Bioinformatics, 2005. **21**(16): p. 3433-4.

167. Dosztanyi, Z., et al., *The pairwise energy content estimated from amino acid composition discriminates between folded and intrinsically unstructured proteins*. J Mol Biol, 2005. **347**(4): p. 827-39.
168. Romero, Obradovic, and K. Dunker, *Sequence Data Analysis for Long Disordered Regions Prediction in the Calcineurin Family*. Genome Inform Ser Workshop Genome Inform, 1997. **8**: p. 110-124.
169. O'Sullivan, J.M., et al., *Gene loops juxtapose promoters and terminators in yeast*. Nat Genet, 2004. **36**(9): p. 1014-8.
170. Ansari, A. and M. Hampsey, *A role for the CPF 3'-end processing machinery in RNAP II-dependent gene looping*. Genes Dev, 2005. **19**(24): p. 2969-78.
171. Zeller, K.I., et al., *Global mapping of c-Myc binding sites and target gene networks in human B cells*. Proc Natl Acad Sci U S A, 2006. **103**(47): p. 17834-9.
172. Sabo, A., et al., *Selective transcriptional regulation by Myc in cellular growth control and lymphomagenesis*. Nature, 2014. **511**(7510): p. 488-492.